Zbornik 24. mednarodne multikonference

# INFORMACIJSKA DRUŽBA

Zvezek B

Proceedings of the 24th International Multiconference

# INFORMATION SOCIETY

Volume B

# IS 2021

## Kognitivna znanost

## Cognitive Science

Uredniki • Editors:
Toma Strle, Borut Trpin, Maša Rebernik, Olga Markič

7. oktober 2021 Ljubljana, Slovenija • 7 October 2021 Ljubljana, Slovenia • http://is.ijs.si

Zbornik 24. mednarodne multikonference

# INFORMACIJSKA DRUŽBA – IS 2021

Zvezek B

Proceedings of the 24th International Multiconference

# INFORMATION SOCIETY – IS 2021

Volume B

## Kognitivna znanost
## Cognitive Science

Uredniki / Editors

Toma Strle, Borut Trpin, Maša Rebernik, Olga Markič

http://is.ijs.si

**7. oktober 2021 / 7 October 2021**
**Ljubljana, Slovenia**

Uredniki:

Toma Strle
Center za Kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Borut Trpin
Filozofska fakulteta, Univerza v Ljubljani

Maša Rebernik
Center za Kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Olga Markič
Filozofska fakulteta, Univerza v Ljubljani

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2021

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 85857795
ISBN 978-961-264-216-7 (PDF)

# PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2021

Štiriindvajseta multikonferenca *Informacijska družba* je preživela probleme zaradi korone v 2020. Odziv se povečuje, v 2021 imamo enajst konferenc, a pravo upanje je za 2022, ko naj bi dovolj velika precepljenost končno omogočila normalno delovanje. Tudi v 2021 gre zahvala za skoraj normalno delovanje konference tistim predsednikom konferenc, ki so kljub prvi pandemiji modernega sveta pogumno obdržali visok strokovni nivo.

Stagnacija določenih aktivnosti v 2020 in 2021 pa skoraj v ničemer ni omejila neverjetne rasti IKTja, informacijske družbe, umetne inteligence in znanosti nasploh, ampak nasprotno – rast znanja, računalništva in umetne inteligence se nadaljuje z že kar običajno nesluteno hitrostjo. Po drugi strani se je pospešil razpad družbenih vrednot, zaupanje v znanost in razvoj. Se pa zavedanje večine ljudi, da je potrebno podpreti stroko, čedalje bolj krepi, kar je bistvena sprememba glede na 2020.

Letos smo v multikonferenco povezali enajst odličnih neodvisnih konferenc. Zajema okoli 170 večinoma spletnih predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic ter 400 obiskovalcev. Prireditev so spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad – seveda večinoma preko spleta. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica (http://www.informatica.si/), ki se ponaša s 45-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2021 sestavljajo naslednje samostojne konference:

- Slovenska konferenca o umetni inteligenci
- Odkrivanje znanja in podatkovna skladišča
- Kognitivna znanost
- Ljudje in okolje
- 50-letnica poučevanja računalništva v slovenskih srednjih šolah
- Delavnica projekta Batman
- Delavnica projekta Insieme Interreg
- Delavnica projekta Urbanite
- Študentska konferenca o računalniškem raziskovanju 2021
- Mednarodna konferenca o prenosu tehnologij
- Vzgoja in izobraževanje v informacijski družbi

Soorganizatorji in podporniki multikonference so različne raziskovalne institucije in združenja, med njimi ACM Slovenija, SLAIS, DKZ in druga slovenska nacionalna akademija, Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

S podelitvijo nagrad, še posebej z nagrado Michie-Turing, se avtonomna stroka s področja opredeli do najbolj izstopajočih dosežkov. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe je prejel prof. dr. Jernej Kozak. Priznanje za dosežek leta pripada ekipi Odseka za inteligentne sisteme Instituta "Jožef Stefan" za osvojeno drugo mesto na tekmovanju XPrize Pandemic Response Challenge za iskanje najboljših ukrepov proti koroni. »Informacijsko limono« za najmanj primerno informacijsko potezo je prejela trditev, da je aplikacija za sledenje stikom problematična za zasebnost, »informacijsko jagodo« kot najboljšo potezo pa COVID-19 Sledilnik, tj. sistem za zbiranje podatkov o koroni. Čestitke nagrajencem!

Mojca Ciglarič, predsednik programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

# FOREWORD - INFORMATION SOCIETY 2021

The 24th *Information Society Multiconference* survived the COVID-19 problems. In 2021, there are eleven conferences with a growing trend and real hopes that 2022 will be better due to successful vaccination. The multiconference survived due to the conference chairs who bravely decided to continue with their conferences despite the first pandemic in the modern era.

The COVID-19 pandemic did not decrease the growth of ICT, information society, artificial intelligence and science overall, quite on the contrary – the progress of computers, knowledge and artificial intelligence continued with the fascinating growth rate. However, COVID-19 did increase the downfall of societal norms, trust in science and progress. On the other hand, the awareness of the majority, that science and development are the only perspectives for a prosperous future, substantially grows.

The Multiconference is running parallel sessions with 170 presentations of scientific papers at eleven conferences, many round tables, workshops and award ceremonies, and 400 attendees. Selected papers will be published in the Informatica journal with its 45-years tradition of excellent research publishing.

The Information Society 2021 Multiconference consists of the following conferences:

- Slovenian Conference on Artificial Intelligence
- Data Mining and Data Warehouses
- Cognitive Science
- People and Environment
- 50-years of High-school Computer Education in Slovenia
- Batman Project Workshop
- Insieme Interreg Project Workshop
- URBANITE Project Workshop
- Student Computer Science Research Conference 2021
- International Conference of Transfer of Technologies
- Education in Information Society

The multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM, SLAIS, DKZ and the second national academy, the Slovenian Engineering Academy. In the name of the conference organizers, we thank all the societies and institutions, and particularly all the participants for their valuable contribution and their interest in this event, and the reviewers for their thorough reviews.

The award for lifelong outstanding contributions is presented in memory of Donald Michie and Alan Turing. The Michie-Turing award was given to Prof. Dr. Jernej Kozak for his lifelong outstanding contribution to the development and promotion of the information society in our country. In addition, the yearly recognition for current achievements was awarded to the team from the Department of Intelligent systems, Jožef Stefan Institute for the second place at the XPrize Pandemic Response Challenge for proposing best counter-measures against COVID-19. The information lemon goes to the claim that the mobile application for tracking COVID-19 contacts will harm information privacy. The information strawberry as the best information service last year went to COVID-19 Sledilnik, a program to regularly report all data related to COVID-19 in Slovenia. Congratulations!

Mojca Ciglarič, Programme Committee Chair
Matjaž Gams, Organizing Committee Chair

# KONFERENČNI ODBORI
# CONFERENCE COMMITTEES

### *International Programme Committee*

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, South Korea
Howie Firth, UK
Olga Fomichova, Russia
Vladimir Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Israel
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Australia
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Mike Bain, Australia
Michela Milano, Italy
Derong Liu, Chicago, USA
Toby Walsh, Australia
Sergio Campos-Cordobes, Spain
Shabnam Farahmand, Finland
Sergio Crovella, Italy

### *Organizing Committee*

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Mitja Lasič
Blaž Mahnič
Klara Vulikić

### *Programme Committee*

Mojca Ciglarič, chair
Bojan Orel,
Franc Solina,
Viljan Mahnič,
Cene Bavec,
Tomaž Kalin,
Jozsef Györkös,
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Mitja Luštrek
Marko Grobelnik
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Boštjan Vilfan
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah
Niko Zimic
Rok Piltaver
Toma Strle
Tine Kolenik
Franci Pivec
Uroš Rajkovič
Borut Batagelj
Tomaž Ogrin
Aleš Ude
Bojan Blažica
Matjaž Kljun
Robert Blatnik
Erik Dovgan
Špela Stres
Anton Gradišek

# KAZALO / TABLE OF CONTENTS

Zbornik 24. mednarodne multikonference

# INFORMACIJSKA DRUŽBA – IS 2021

Zvezek B

Proceedings of the 24th International Multiconference

# INFORMATION SOCIETY – IS 2021

Volume B

## Kognitivna znanost
## Cognitive Science

Uredniki / Editors

Toma Strle, Borut Trpin, Maša Rebernik, Olga Markič

http://is.ijs.si

**7. oktober 2021 / 7 October 2021**
**Ljubljana, Slovenia**

## PREDGOVOR

Na letošnji konferenci Kognitivna znanost sodelujejo avtorice in avtorji z različnih disciplinarnih področij in predstavljajo tako empirične rezultate svojih raziskav kot tudi teoretska raziskovanja z najrazličnejših področij – od psihologije in jezikoslovja do nevrofenomenologije, filozofije in umetne inteligence.

Upamo, da bo letošnja disciplinarno in metodološko bogata konferenca odprla prostor za izmenjavo zanimivih raziskovalnih idej ter povezala znanstvenice in znanstvenike z različnih disciplinarnih področij, ki se ukvarjajo z vprašanji kognicije.

Toma Strle
Borut Trpin
Maša Rebernik
Olga Markič

## FOREWORD

At this year’s Cognitive Science conference, the authors present their empirical studies as well as theoretical research from a diverse range of disciplinary backgrounds – from psychology and linguistics to neurophenomenology, philosophy, and artificial intelligence.

We hope that this year's cognitive science conference – rich in disciplinary approaches and methodologies – will open space for exchanging intriguing research ideas and will bring together scientists from a diverse range of areas related to the exploration of the human mind.

Toma Strle
Borut Trpin
Maša Rebernik
Olga Markič

**PROGRAMSKI ODBOR / PROGRAMME COMMITTEE**

Toma Strle, Center za Kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Borut Trpin, Filozofska fakulteta, Univerza v Ljubljani

Maša Rebernik, Center za Kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Olga Markič, Filozofska fakulteta, Univerza v Ljubljani

Urban Kordeš, Center za kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Matjaž Gams, Odsek za inteligentne sisteme, Institut »Jožef Stefan«, Ljubljana

**ORGANIZACIJSKI ODBOR / ORGANIZING COMMITTEE**

Toma Strle, Center za kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Borut Trpin, Filozofska fakulteta, Univerza v Ljubljani

Maša Rebernik, Center za Kognitivno znanost, Pedagoška fakulteta, Univerza v Ljubljani

Olga Markič, Filozofska fakulteta, Univerza v Ljubljani

# Nevrofenomenološka študija skupinskih dinamik v spletnem učnem okolju: Preliminarni rezultati

## Neurophenomenological Study of Group Dynamics in the Online Learning Environment: Preliminary results


Jaša Černe
Center za kognitivno znanost
Univerza v Ljubljani
Ljubljana, Slovenija
jasa.cerne@pef.uni-lj.si

Selma Berbić
MEi:CogSci
Univerza v Ljubljani
Ljubljana, Slovenija
selmaberbic2@gmail.com

Mateja Kalan
MEi:CogSci
Univerza v Ljubljani
Ljubljana, Slovenija
mateja.kalan07@gmail.com

Lucija Mihić Zidar
MEi:CogSci
Univerza v Ljubljani
Ljubljana, Slovenija
lucijamihiczidar@gmail.com

Uršek Slivšek
MEi:CogSci
Univerza v Ljubljani
Ljubljana, Slovenija
slivsek@protonmail.com

Urban Kordeš
Center za kognitivno znanost
Univerza v Ljubljani
Ljubljana, Slovenija
urban.kordes@pef.uni-lj.si



## POVZETEK

Učno okolje je prostor, v katerem se med udeleženimi v učnem procesu ustvarjajo kompleksne skupinske dinamike. V prispevku predstavimo preliminarne rezultate eksploratorne nevrofenomenološke študije, v kateri smo preučevali takšne dinamike v spletnem učnem okolju. Udeleženci so na štirih srečanjih merili elektrodermalno aktivnost in ob naključnih trenutkih vzorčili doživljanje. Po vsakem srečanju so izvajali fenomenološke intervjuje in se spoznavali s podatki. Rezultati so pokazali obstoj različnih skupinskih dinamik na ravni doživljanja in psihofiziologije, kar predstavlja osnovo za nadaljnjo nevrofenomenološko analizo. Nadejamo se, da bodo ugotovitve ponudile svež uvid v vedno pogostejše spletno poučevanje in pomagale oblikovati boljše učne pristope.



## KLJUČNE BESEDE

Skupinska dinamika, nevrofenomenologija, vzorčenje doživljanja, elektrodermalna aktivnost, fiziološka sinhronizacija, spletno učno okolje



## ABSTRACT

A learning environment is a space wherein complex group dynamics form between those who participate in the learning process. In this paper, we present the preliminary results of a exploratory neurophenomenological study in which we examined such dynamics in an online learning environment. Throughout four sessions, participants measured electrodermal activity and sampled their experience at random moments. After each session, they conducted phenomenological interviews and familiarized themselves with the data. The results showed the existence of various group dynamics at the level of experience and psychophysiology, which represents the basis for further neurophenomenological analysis. We hope that the findings will offer fresh insight into the increasingly common online teaching and help shape better learning approaches.



## KEYWORDS

Group dynamics, neurophenomenology, experience sampling, electrodermal activity, physiological synchrony, online learning environment




*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*



## 1 UVOD

Učno okolje sestavljajo učitelji in učenci, ki sodelujejo v izmenjavi znanja. Čeprav gre v osnovi za delovanje avtonomnih posameznikov, postane to delovanje včasih zelo usklajeno, tj. tvorijo se skupinske dinamike [1]. V zadnjem času se je zvrstilo več študij, ki skušajo raziskati naravo tovrstnih dinamik z družnim raziskovanjem doživljanja (prvoosebni vidik) in nevrološke aktivnosti (tretjeosebni vidik) [2, 3, 4, 5, 6]. Pokazale so, da obstaja korelacija med kolektivnim doživljajskim stanjem učencev v razredu (npr. čustveno atmosfero) in pripadajočo nevrološko oziroma psihofiziološko sinhronizacijo [2, 3, 4]. Kljub temu, da se poučevanje vztrajno širi na splet [7], kar lahko predrugači običajne skupinske dinamike [8], se nobena takšna študija še ni ukvarjala s spletnim učnim okoljem. Z raziskavo, ki jo opišemo v tem prispevku, smo želeli zapolniti to vrzel.

Sodobni kognitivni znanosti povezovanje doživljajskega in nevrološkega nivoja ni tuje [9, 10]. Tretjeosebne opise, ki jih podaja npr. nevroznanost, je potrebno osmisliti skozi prizmo pripadajočih prvoosebnih opisov [11]. Toda slednji so pogosto pridobljeni s tehnikami, ki dajejo prednost posploševanju in formalizaciji, zapostavljajo pa veljavnost in ločljivost [11, 12]. Zaradi tega lahko ostane ogromno nevroloških variabilnosti, kot tudi morebitnih korelacij med prvoosebnim in tretjeosebnim nivojem, spregledanih [13, 14]. Potencialno rešitev je v svojem nevrofenomenološkem programu predlagal Francisco Varela

[11]. Poudaril je pomembnost poglobljenega, a sistematičnega pridobivanja prvoosebnih podatkov in združevanja prvoosebnega in tretjeosebnega nivoja po principu vzajemnega omejevanja. Več študij je pokazalo, da takšno nevrofenomenološko raziskovanje ni samo izvedljivo, pač pa lahko ponudi svež uvid v pereče probleme kognitivnih znanosti (za nedavni pregled glej [12]). Tehnika za pridobivanje prvoosebnih podatkov, ki je bila že večkrat uspešno uporabljena v nevrofenomenološkem kontekstu [15, 17], je opisno vzorčenje izkustva (OVI) [18, 19]. Sestavni del tehnike OVI sta naključno vzorčenje doživljanja in kasnejši fenomenološki intervjuji, pri čemer sta tako spraševanje kot tudi poročanje o doživljanju smatrana za spretnosti, v katerih se je potrebno uriti [19].

Za razumevanje nevrološke podstati doživljajskih stanj se pogosto uporabljajo mere delovanja avtonomnega živčnega sistema (AŽS), kot je npr. elektrodermalna aktivnost (EDA) [20], [21, 22]. EDA je produkt interakcije lokalnih procesov v koži in delovanja simpatičnega dela AŽS ter se navadno uporablja kot indikator vzburjenosti, čustev in stresa [24, 25]. Različne mere sinhronizacije EDA med več udeleženci so se nedavno uveljavile kot učinkovit pokazatelj skupinskih dinamik, povezanih npr. z empatijo [26], s povezanostjo med govorniki in občinstvom [27] ter s povečano slušno osredotočenostjo [28]; pa tudi skupinskih dinamik, ki se oblikujejo v učnem okolju, npr. nižja vključenost v učni proces [29], mentalni napori skupine [30] in čustvena atmosfera [27]. Kljub obetavnim rezultatom pa doslej še ni bilo opravljene študije, ki bi mero EDA na nevrofenomenološki način združila s sodobno metodo za pridobivanje prvoosebnih podatkov, kot je npr. tehnika OVI.

V nadaljevanju predstavimo preliminarne rezultate eksploratorne nevrofenomenološke raziskave, v kateri smo na ekološko veljaven način preučevali doživljanje in EDA udeležencev v spletnem učnem okolju. Odgovoriti smo želeli na štiri raziskovalna vprašanja: (RV1) Kaj doživljajo študenti in izvajalci tekom spletnih predavanj? (RV2) Ali lahko ob istih časovnih trenutkih prepoznamo skupinske dinamike na doživljajskem nivoju? (RV3) Ali se med udeleženimi v učnem procesu tekom spletnih predavanj pojavljajo skupinske dinamike oziroma sinhronizacije na nivoju EDA? (RV4) Ali obstajajo povezave med doživljanjem in EDA udeleženih v učnem procesu?

# 2 METODA

## 2.1 Oris raziskave

Raziskava je vključevala štiri spletna srečanja (pilotno in tri raziskovalna) v okviru predavanj na skupnem Interdisciplinarnem srednjeevropskem magistrskem študijskem programu Kognitivna znanost (MEi:CogSci). Sodelovanje v raziskavi je bilo izrazito aktivno oziroma participatorno. Med srečanjem so udeleženci vzorčili doživljanje in merili EDA, po srečanju pa so opravili fenomenološke intervjuje o izbranih vzorcih in krajšo sprotno analizo. Fazi zbiranja podatkov je sledila obširnejša analiza, v načrtu pa imamo opraviti še nevrofenomenološko analizo, v kateri bo izveden poskus integracije prvoosebnih in tretjeosebnih podatkov. Splošno shemo poteka raziskave prikazuje Slika 1.


**Slika 1: Shema poteka raziskave**

## 2.2 Udeleženci

V raziskavi je sodelovalo petnajst udeležencev (enajst žensk; povprečna starost = 27,0 let; *SD* = 7,4) od tega štirinajst študentov in en izvajalec. Izvajalec je imel večletne izkušnje z raziskovanjem doživljanja, študenti pa so pred raziskavo opravili trening vzorčenja doživljanja in izvajanja fenomenoloških intervjujev. Po vzoru tehnike OVI [19] je vsak študent vzorčil doživljanje vsaj 9 dni, pridobil vsaj 39 vzorcev, bil intervjuvan o vsaj 15 svojih vzorcih in opravil intervju o vsaj 15 vzorcih nekoga drugega. Pred prvim srečanjem so bili udeleženci seznanjeni z raziskavo, pridobljeno pa je bilo tudi njihovo soglasje za sodelovanje. Udeleženci so lahko s sodelovanjem v raziskavi opravili del obveznosti pri študiju.

## 2.3 Pripomočki in tehnike

Za merjenje EDA je bil uporabljen brezžični nadlahtni merilnik BodyMedia SenseWear. Merilnik je beležil EDA štirikrat na minuto in shranjeval podatke v interni spomin.

Prvoosebni podatki so bili pridobljeni s tehniko vzorčenja doživljanja, osnovano na tehniki OVI [19]. Signal za vzorčenje je sprožila aplikacija, naključno v intervalu od 5 do 15 minut. Za vzorčenje je bil uporabljen vprašalnik, ki se je delno razlikoval med pilotnim in ostalimi srečanji. Na pilotnem so udeleženci poročali o kontekstu in doživljanju v zadnjem trenutku pred signalom za vzorčenje, podali pa so lahko tudi komentar in opazke o doživljanju pred tem. Na vseh ostalih srečanjih so udeleženci poročali o istih postavkah kot na pilotnem srečanju in dodatno o doživljanju, ki je bilo v zadnjem trenutku pred signalom za vzorčenje v ospredju, podali pa so tudi odgovor na dve vprašanji z vnaprej predvidenimi odgovori. Pri prvem so označili stopnjo, do katere so bili v trenutku vzorčenja vpeti v vsebino predavanja (označili so lahko: *aktivna vpetost*, *vpetost*, *delna vpetost*, *delna odsotnost*, *odsotnost* ali *drugo*), pri drugem pa vrsto socialnega doživljanja, ki je bila takrat prisotna (označili so lahko: *brez socialnega doživljanja*, *preverjanje doživljanja drugih*, *občutek kolektivnega doživljanja*, *socialno uravnavanje* ali *drugo*).

Doživljajski vzorci so bili razširjeni in preverjeni s tehniko fenomenološkega intervjuja, osnovano delno na ekspozicijskem [26] in delno na mikrofenomenološkem [31] intervjuju.

## 2.4 Postopek

Vsa srečanja so potekala na spletni platformi Zoom. Pilotno srečanje je bilo namenjeno spoznavanju protokola raziskave in raziskovanega pojava, testiranju uporabljene tehnologije ter natančni specifikaciji raziskovalnih vprašanj. Na podlagi podatkov, pridobljenih na pilotnem srečanju, je bil oblikovan vprašalnik za vzorčenje doživljanja.

Na začetku vsakega srečanja so si udeleženci namestili merilnik za merjenje EDA, sledili sta dve minuti mirovanja, nato se je začelo predavanje. Tekom predavanja se je od pet do šestkrat predvajal zvočni signal, po katerem so imeli udeleženci na voljo eno do dve minuti za vzorčenje doživljanja. Po srečanju so udeleženci zbrane podatke naložili na spletni repozitorij.

Študenti so v času do tri dni po vsakem srečanju izvedli sprotno analizo prvoosebnih in tretjeosebnih podatkov, med tremi do šestimi dnevi po srečanju pa še fenomenološke intervjuje o izbranih doživljajskih vzorcih. O vsakem intervjuju so zapisali kratko poročilo.

## 2.5 Analiza

Analizo podatkov smo izvajali med in po koncu zbiranja podatkov. Glavni cilj analize je bil prepoznavanje vzorcev, ki namigujejo na obstoj skupinskih dinamik.

***Sprotna analiza.*** Sprotna analiza podatkov EDA je vključevala vizualno identifikacijo sinhronizacij v signalih, sprotna analiza doživljajskih podatkov pa primerjavo vzorcev in preliminarno kategorizacijo. Izsledki sprotnih analiz so informirali nadaljnje faze raziskovanja in analize.

***Doživljajska analiza.*** Primarne podatke za doživljajsko analizo so predstavljali odgovori na odprto vprašanje o doživljanju v zadnjem trenutku pred signalom za vzorčenje, odgovori na ostale postavke vprašalnika in poročila o intervjujih pa so služili dodatnemu preverjanju. Analiza je potekala po vzoru smernic za doživljajsko [32] in kvalitativno analizo [33, 34]. Najprej smo označili »satelitske« [31] dimenzije doživljanja, nato pa z induktivnim pristopom odprtega kodiranja [33] vsakemu vzorcu pripisali kategorije prvega reda. S primerjalno analizo smo prvotne kategorije po potrebi prilagodili, oblikovali višjenivojske kategorije in dobljene kategorije definirali. Na koncu smo izbrali tiste kategorije, ki so bile najpogostejše in/ali najbolj relevantne z vidika zastavljenih raziskovalnih vprašanj.

***Analiza EDA.*** Analiza EDA je vključevala izračun sinhronizacij med pari udeležencev (od tu naprej parnih sinhronizacij) in izračun povprečnih parnih sinhronizacij (PPS) različnih skupin: (1) skupin vsaj treh med seboj sinhroniziranih udeležencev ($r \geq 0{,}40$)[1]; (2) vnaprej definiranih skupin (vsi udeleženci; samo študenti; izvajalec z vsakim študentom).

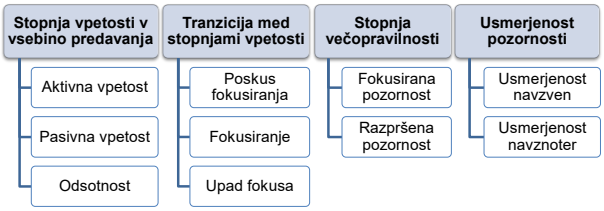
Petminutne odseke signalov EDA[2], ki so bili posneti v času pred vzorčenjem doživljanja, smo ročno pregledali in odstranili takšne, ki so vsebovali artefakte [24]. Pred nadaljnjo analizo smo dobljene signale standardizirali. Za izračun parnih sinhronizacij smo uporabili prilagojen algoritem Marci in Orra [26]. Sinhronizacijo EDA enega para pri enem vzorčenju smo izračunali kot povprečje dvanajstih Pearsonovih korelacij, pridobljenih s pomikanjem tekočega okna dolžine osmih meritev (dvominutni odsek) po eno meritev naprej, dokler nismo obdelali vseh dvajsetih meritev (petminutni odsek).

# 3 REZULTATI

Cilj raziskave je bil opisati doživljajsko pokrajino udeležencev med spletnimi predavanji (RV1) in preveriti, ali se na doživljajskem (RV2) in psihofiziološkem (RV3) nivoju, ter na obeh nivojih skupaj (RV4), porajajo skupinske dinamike. V nadaljevanju predstavimo preliminarne rezultate, ki se navezujejo na RV1, RV2 in RV3.

## 3.1 Doživljanje udeležencev (RV1)

Kot je razvidno iz Slike 2, je doživljajska analiza pokazala, da lahko doživljanje udeležencev (izvajalca in študentov) tekom spletnih predavanj opišemo s štirimi krovnimi kategorijami, vezanimi na osredotočenost in socialno doživljanje.




**Slika 2: Hierarhija izbranih doživljajskih kategorij**

***Stopnja vpetosti v vsebino predavanja.*** Nekateri študenti so se v trenutku vzorčenja aktivno ukvarjali z relevantno vsebino ali pa so kako drugače izkazovali zanimanje zanjo; poročali so npr. o vizualizaciji in interpretaciji relevantnih konceptov, povezovanju z obstoječim znanjem, pa tudi o pričakovanju sledeče vsebine. Tako je zapisala Mara[3]: "Slušno zaznavam besede [izvajalca], subtilno si predstavljam nadaljnji potek predavanja, kot ga opisuje, na način, da interpretiram pomen besed v nesimbolnih mislih." Izvajalec je sicer zmeraj aktivno posredoval vsebino, a je včasih vseeno poročal o večjem zanimanju. Takšne primere smo imenovali *aktivna vpetost* ($n$ = 88). Včasih so študenti vsebino predavanja sicer zaznavali, a ne tako pozorno in z njo niso ničesar aktivno počeli. Tudi izvajalec je včasih poročal o manjši zbranosti ali naveličanosti. Takšne primere smo uvrstili v podkategorijo *pasivna vpetost* ($n$ = 50). Nazadnje smo prepoznali tudi več primerov *odsotnosti* ($n$ = 30), ko v doživljajskih pokrajinah študentov ni bilo mogoče zaznati vsebine predavanja, izvajalec pa je poročal npr. o zmedenosti.

***Tranzicija med stopnjami vpetosti.*** Doživljanje udeležencev se je včasih nanašalo na prehodne faze med *stopnjami vpetosti v vsebino predavanja*. Nekateri udeleženci so v trenutku vzorčenja poročali o *poskusu fokusiranja* ($n$ = 19) oziroma prizadevanju za aktivnejšo vpetost v vsebino predavanja. Mara je na primer zapisala: "Doživljam težnjo po poglobitvi pozornosti na predavanje." Drugi so težnjo po fokusiranju že začeli udejanjati – signal za vzorčenje jih je ujel v procesu *fokusiranja* ($n$ = 15), ko so pozornost že preusmerjali na vsebino predavanja. Spet drugi so poročali o pravkaršnjemu *upadu fokusa* ($n$ =19), bodisi zaradi utrujenosti, zaspanosti, lakote ali naveličanosti.

[1] Kriterij $r \geq 0{,}40$ razumemo kot spodnjo mejo srednje močne korelacije [23].

[2] Doživljajski podatki so bili omejeni izključno na zadnji trenutek pred signalom za vzorčenje, zato v analizi EDA nismo upoštevali celih signalov, ampak zgolj petminutne odseke, ki so bili posneti pred vzorčenjem doživljanja.

[3] Izseki, ki jih podajamo ob opisih kategorij, so urejeni tako, da ne razkrivajo identitet udeležencev in so po potrebi osnovno lektorirani.

***Stopnja večopravilnosti.*** Doživljanje udeležencev je bilo mogoče razdeliti tudi glede na številčnost aktivnosti, na katere so bili pozorni. Včasih so bili osredotočeni le na vsebino predavanja – takšne primere smo imenovali *fokusirana pozornost* ($n$ = 15). Med njimi najdemo zapis Mare: "Sem v stanju pričakovanja, občutim radovednost kot željo po razjasnitvi pojma izomorfizem." Občasno so bili udeleženci, npr. Zoja, poleg predavanja osredotočeni še na kaj drugega: "Poslušam in zdi se mi (čutim), da vem, o čem predavatelj govori [...]. Moja pozornost je sicer rahlo razpršena – misli mi tavajo na več koncev, predvsem preverjam, kaj vse moram še danes narediti." Takšne primere smo označili z *razpršeno pozornostjo* ($n$ = 16).

***Usmerjenost pozornosti.*** Doživljanja udeležencev so včasih zaznamovali občutki, vezani na druge (virtualno) prisotne na srečanju; udeleženci so se zavedali drugih, skušali so ugotoviti, kaj drugi doživljajo, ali pa so jih opazovali na Zoomu. Te primere smo združili v podkategorijo *usmerjenost navzven* ($n$ = 44). Toda socialnega doživljanja ni bilo zmeraj zaznati; včasih so udeleženci opazovali svoje doživljanje, izvajali samorefleksijo, ali pa se samoopazovali na Zoomu. Takšne zapise smo označili z *usmerjenostjo navznoter* ($n$ = 27).

## 3.2 Doživljajske skupinske dinamike (RV2)

Skupinsko dinamiko na doživljajskem nivoju smo definirali kot skupino treh ali več udeležencev, katerih istočasno podane vzorce doživljanja smo uvrstili v isto podkategorijo (glej Sliko 2). Skupno smo prepoznali 56 primerov skupinskih dinamik, od tega 19 za prvo, 19 za drugo in 18 za tretje srečanje. 40-krat so skupinske dinamike tvorili študenti, 16-krat pa študenti in izvajalec. Najpogosteje so bile skupinske dinamike vezane na podkategorijo *aktivna vpetost* ($n$ = 18). Najbolj opazno usklajenost smo prepoznali pri petem vzorčenju tretjega srečanja, ko so tako izvajalec kot sedem študentov sočasno poročali o *aktivni vpetosti*. Izvajalec je takrat zapisal: "Stanje zaganjanja v predavateljski tok – ne še čisto tam. Tokrat je nemir v ozadju močnejši, je pa tudi višja energija – bolj aktivno 'sodelujem' pri oblikovanju predavanja." Ena izmed študentk, Ajša, pa je poročala: "Zanimanje za to, kar [izvajalec] govori, kar sem čutila kot željo, da si o tem kaj napišem ter da slišim vse, kar izreče, da ne izgubim toka govora."

Skupinske dinamike so se tekom vzorčenj posameznega srečanja sistematično spreminjale. Denimo na prvem srečanju smo pri četrtem vzorčenju zaznali splošen upad osredotočenosti tako pri izvajalcu kot pri študentih. Do tretjega vzorčenja so izvajalec in večina študentov ($M$ = 9,3; $SD$ = 2,3) poročali o *aktivni vpetosti*, manj študentov pa je poročalo o *pasivni vpetosti* ($M$ = 2,7; $SD$ = 2,4) in *odsotnosti* ($n$ = 1). Zatem izvajalec ni več poročal o *aktivni vpetosti*, prav tako je o njej poročalo bistveno manj študentov ($M$ = 5,0; $SD$ = 0,0), število tistih, ki so bili *pasivno vpeti* ($M$ = 5,0; $SD$ = 2,0) v vsebino predavanja, ali so bili *odsotni* ($M$ = 3,0; $SD$ = 0,0), pa se je dvignila. Izvajalec je takrat zapisal: "Čutim se odsotnega, avtomatično govorjenje – tema mi je dolgočasna, rad bi, da jo čim prej zrecitiram, da grem naprej na bolj zanimivo vsebino."

## 3.3 Psihofiziološke skupinske dinamike (RV3)

Analiza podatkov EDA je pokazala skupno 25 skupin s tremi ali več medsebojno parno sinhroniziranimi člani. Na prvem srečanju (pet vzorčenj) smo prepoznali šest sinhroniziranih skupin, na drugem (šest vzorčenj) deset in na tretjem (pet vzorčenj) devet. Sinhronizirane skupine se niso ohranjale prek več vzorčenj enega srečanja. Najvišja PPS je znašala 0,78 (tretje vzorčenje tretjega srečanja), povprečje PPS vseh skupin pa je bilo 0,62 ($SD$ = 0,08).

Pri vnaprej definiranih skupinah smo največjo skupinsko dinamiko opazili na prvem srečanju, kjer je bila PPS vseh udeležencev 0,20 ($SD$ = 0,54), vseh študentov 0,14 ($SD$ = 0,53), izvajalca s študenti pa 0,40 ($SD$ = 0,58). Pri drugem vzorčenju je bila PPS vseh udeležencev 0,19 ($SD$ = 0,29), vseh študentov 0,17 ($SD$ = 0,30), izvajalca s študenti pa 0,29 ($SD$ = 0,22). Pri zadnjih treh vzorčenjih se je PPS gibala okrog 0. Na drugem srečanju smo prepoznali manj očitne skupinske dinamike. Pri prvem vzorčenju je PPS izvajalca s študenti znašala 0,17 ($SD$ = 0,55), pri drugem 0,15 ($SD$ = 0,38) in pri šestem prav tako 0,15 ($SD$ = 0,29). Pri četrtem vzorčenju je znašala PPS vseh udeležencev 0,12 ($SD$ = 0,37), vseh študentov pa 0,16 ($SD$ = 0, 38). Sicer se je PPS gibala okrog 0. Na tretjem srečanju nismo prepoznali PPS večjih od 0. Za vsa tri srečanja je povprečje PPS vseh udeležencev znašalo 0,04 ($SD$ = 0,07), vseh študentov 0,03 ($SD$ = 0,10) in izvajalca s študenti 0,05 ($SD$ = 0,15).

# 4 DISKUSIJA

V prispevku smo pokazali, da se tudi v spletnem učnem okolju, kjer udeleženci niso fizično prisotni, tvorijo doživljajski in psihofiziološki vzorci koordiniranega delovanja tako med študenti kot med študenti in izvajalcem. Da bi videli, ali se prepoznane skupinske dinamike porajajo na obeh nivojih hkrati, bomo v naslednjem koraku izvedli nevrofenomenološko analizo, v kateri bomo izsledke neodvisne doživljajske analize preverili z dodatno analizo EDA in izsledke neodvisne analize EDA z dodatno doživljajsko analizo. Upamo, da bodo končni rezultati poglobili razumevanje skupinskih dinamik, ki se tvorijo v spletnem učnem okolju. Ker so določene skupinske dinamike povezane z akademsko uspešnostjo [1, 35, 36], upamo, da bodo naši rezultati pripomogli tudi k izboljšanju učnih pristopov.

Določene pomanjkljivosti raziskave najdemo v načinu izvedbe, uporabljeni tehnologiji in izbrani metodi. Prvič, dejstvo, da je bila raziskava izvedena v naravnem okolju je po eni strani povečalo njeno ekološko veljavnost, po drugi strani pa otežilo posploševanje zaradi nezmožnosti zagotavljanja univerzalnosti eksperimentalnega okolja. Drugič, merilnik, s katerim smo pridobivali podatke EDA, je namenjen za uporabo na nadlahti, ki je optimalna lokacija z vidika nizke invazivnosti, ne pa tudi z vidika pridobivanja podrobnih podatkov o psihofiziološkem stanju uporabnika [24, 37]. Tretjič, podatke o psihofiziologiji smo pridobivali zgolj s pomočjo mere EDA, medtem ko bi lahko kombinirana uporaba več senzorjev psihofiziologije omogočila podrobnejši uvid v delovanje AŽS [38]. Četrtič, doživljajski vzorci so bili mestoma premalo natančni, fenomenološki intervjuji, s katerimi smo reševali ta problem, pa so bili opravljeni le o izbranih vzorcih in včasih šele tretji dan po srečanju, kar je otežilo priklic informacij iz spomina. Izvajanje intervjujev o vseh vzorcih v krajšem času od vzorčenja bi po drugi strani bistveno povečalo že tako visoke zahteve, ki jih je raziskava polagala na pleča udeležencev.

Metodološki izziv za prihodnje raziskave je torej najti način, kako sočasno zagotoviti visoko ekološko veljavnost in univerzalnost okoljskih dejavnikov, kako sočasno zadovoljiti potrebo po nizki invazivnosti in visoki odzivnosti merilnikov

psihofiziologije, ter kako uskladiti potrebe po pridobivanju podrobnih in veljavnih prvoosebnih podatkov na način, ki ne bo pretirano zahteven za udeležence.

## ZAHVALA

Hvala Gregorju Geršaku za merilnike in uvod v svet psihofiziologije, Tini Giber za praktične nasvete pri rokovanju z merilniki in vsem študentom za potrpežljivo ter vse prej kot pasivno sodelovanje v raziskavi.

## LITERATURA

[1] Maria R. Reyes, Marc A. Brackett, Susan E. Rivers, Mark White in Peter Salovey. 2012. Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology* 104, 3, (Avgust, 2012), 700–712. https://doi.org/10.1037/a0027268

[2] Lauri Ahonen, Benjamin U. Cowley, Arto Hellas in Kai Puolamäki. 2018. Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific reports* 8, 1, (Februar, 2018), 3138. https://doi.org/10.1038/s41598-018-21518-3

[3] Dana Bevilacqua, Ido Davidesco, Lu Wan, Kim Chaloner, Jess Rowland, Mingzhou Ding, David Poeppel in Suzanne Dikker. 2019. Brain-to-Brain Synchrony and Learning Outcomes Vary by Student-Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. *Journal of cognitive neuroscience* 31, 3, (Marec, 2019), 401–411. https://doi.org/10.1162/jocn_a_01274

[4] Anne-Marie Brouwer, Ivo V. Stuldreher in Nattapong Thammasan. 2019. Shared attention reflected in eeg, electrodermal activity and heart rate. *CEUR Workshop Proceedings, 2019 Socio-Affective Technologies: An Interdisciplinary Approach*, (Oktober, 2019), 27-3. http://resolver.tudelft.nl/uuid:d9afe398-c8e5-42ff-ab93-7b9313cf6d2e

[5] Suzanne Dikker, Lu Wan, Ido Davidesco, Lisa Kaggen, Matthias Oostrik, James McClintock, Jess Rowland, Georgios Michalareas, Jay J. Van Bavel, Mingzhou Ding in David Poeppel. 2017. Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current biology: CB* 27, 9, (Maj, 2017), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

[6] Yu Zhang, Fei Qin, Bo Liu, Xuan Qi, Yingying Zhao in Dan Zhang. 2020. Wearable Neurophysiological Recordings in Middle-School Classroom Correlate with Students' Academic Performance. *Frontiers in human neuroscience* 12, 457, (November, 2018). https://doi.org/10.3389/fnhum.2018.00457

[7] Shivangi Dhawan. 2020. Online Learning: A Panacea in the Time of COVID-19 Crisis. *Journal of Educational Technology Systems* 49, 1, (Junij, 2020), 5–2. https://doi.org/10.1177/0047239520934018

[8] Lynn Clouder, Jayne Dalley, Julian Hargreaves, Sally Parkes, Julie Sellars in Jane Toms. 2006. Electronic [re]constitution of groups: Group dynamics from face-to-face to an online setting. *Computer Supported Learning 1*, (December, 2006), 467–480. https://doi.org/10.1007/s11412-006-9002-0

[9] John T. Cacioppo, Louis G. Tassinary in Gary Berntson (ur.). 2007. *Handbook of psychophysiology* (3. Izdaja). Cambridge University press. New York, NY.

[10] Michael Gazzaniga, Richard B. Ivry in George R. Mangun. 2019. *Cognitive neuroscience: The biology of the mind* (5. izdaja). Norton, New York, NY.

[11] Francisco J. Varela. 1996. Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3, 4, (April, 1996), 330-49.

[12] Aviva Berkovich-Ohana, Yair Dor-Ziderman, Fynn-Mathis Trautwein, Yoav Schweitzer, Ohad Nave, Stephen Fulder in Yochai Ataria. 2020. The Hitchhiker's Guide to Neurophenomenology - The Case of Studying Self Boundaries with Meditators. *Frontiers in psychology* 11, 1680, (Julij, 2020). https://doi.org/10.3389/fpsyg.2020.01680

[13] Antoine Lutz, Jean-Philippe Lachaux, Jacques Martinerie, Francisco Varela, 2002. Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *PNAS*, 99, 3, 1586–1591. https://doi.org/10.1073

[14] Claire Petitmengin, Vincent Navarro in Michel Le Van Quyen, 2007. Anticipating seizure: Pre-reflective experience at the center of neuro-phenomenology. *Consciousness and Cognition*, 16, 3, 746–764. https://doi.org/10.1016/j.concog.2007.05.006

[15] Simone Kühn, Charles Fernyhough, Benjamin Alderson-Day in Russell T. Hurlburt. 2014. Inner experience in the scanner: Can high fidelity apprehensions of inner experience be integrated with fMRI? *Frontiers in Psychology* 5, 1393, (December, 2014). https://doi.org/10.3389/fpsyg.2014.01393

[16] Russell T. Hurlburt, Ben Alderson-Day, Charles Fernyhough in Simone Kühn. 2015. What goes on in the resting-state? A qualitative glimpse into resting-state experience in the scanner. *Frontiers in psychology* 6, 1535, (Oktober, 2015). https://doi.org/10.3389/fpsyg.2015.01535

[17] Russell T. Hurlburt, Ben Alderson-Day, Simone Kühn in Charles Fernyhough. 2016. Exploring the ecological validity of thinking on demand: Neural correlates of elicited vs. spontaneously occurring inner speech. PLoS ONE 11, 2, (Februar, 2016), e0147932. https://doi.org/10.1371/journal.pone.0147932

[18] Russell T. Hurlburt in Sarah A. Akhter. 2006. The Descriptive Experience Sampling method. Phenomenology and the Cognitive Sciences 5, (November, 2006), 271–301. https://doi.org/10.1007/s11097-006-9024-0

[19] Russell T. Hurlburt, Sharon Jones-Forrester (Sod.), Michael J. Kane (Sod.), Ricardo Cobo (Sod.), Sarah A. Akhter (Sod.), Chris Heavey (Sod.) in Arva Bensaheb (Sod.). 2011. *Investigating pristine inner experience: Moments of truth*. Cambridge University Press. Cambridge, UK. https://doi.org/10.1017/CBO9780511842627

[20] Antoine Bechara, Hanna Damasio, Daniel Tranel in Antonio R. Damasio. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275, 5304, (Februar, 1997), 1293–1295. https://doi.org/10.1126/science.275.5304.1293

[21] Lea K. Hildebrandt, Cade McCall, Haakon G. Engen in Tania Singer. 2016. Cognitive flexibility, heart rate variability, and resilience predict fine-grained regulation of arousal during prolonged threat. *Psychophysiology* 53, 6, (Junij, 2016), 880–890. https://doi.org/10.1111/psyp.12632

[22] Cade McCall, Lea K. Hildebrandt, Boris Bornemann in Tania Singer. 2015. Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition* 38, (December, 2015), 60–70. https://doi.org/10.1016/j.concog.2015.09.011

[23] Haldun Akoglu. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18, 3, (September, 2018), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001.

[24] Wolfram Boucsein. 2012. *Electrodermal Activity* (2. izdaja). Springer, New York. NY.

[25] Gregor Geršak. 2020. Electrodermal activity - a beginner' s guide. *Elektrotehniški Vestnik* 87, 4, (Januar, 2020), 175–182.

[26] Carl D. Marci in Scott P. Orr. 2006. The effect of emotional distance on psychophysiologic concordance and perceived empathy between patient and interviewer. *Applied Psychophysiology and Biofeedback* 31, 2 (Junij, 2006), 115–128. https://doi.org/10.1007/s10484-006-9008-4

[27] Shkurta Gashi, Elena Di Lascio in Silvia Santini. 2019. Using Unobtrusive Wearable Sensors to Measure the Physiological Synchrony Between Presenters and Audience Members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1, (Marec, 2019), 1–19. https://doi.org/10.1145/3314400

[28] Ivo V. Stuldreher, Nattapong Thammasan, Jan B. F. van Erp in Anne-Marie Brouwer. 2020. Physiological Synchrony in EEG, Electrodermal Activity and Heart Rate Detects Attentionally Relevant Events in Time. *Frintiers in Neuroscience* 14, 575521, (December, 2020), https://doi.org/10.3389/fnins.2020.575521

[29] Elena Di Lascio, Shkurta Gashi in Silvia Santini. 2018. Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3, (September, 2018), 1–21. https://doi.org/10.1145/3264913

[30] Muhterem Dindar, Sanna Järvelä in Eetu Haataja. 2020. What does physiological synchrony reveal about metacognitive experiences and group performance? *British Journal of Educational Technology* 51, (Junij, 2020), 1577–1597. https://doi.org/10.1111/bjet.12981

[31] Claire Petitmengin. 2006. Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive Sciences* 5, 3–4, (November, 2006), 229–269. https://doi.org/10.1007/s11097-006-9022-2

[32] Claire Petitmengin, Anne Remillieux in Camila Valenzuela-Moguillansky. 2019. Discovering the structures of lived experience: *Towards a micro-phenomenological analysis method. Phenomenology and the Cognitive Sciences* 18, 4, (September, 2019), 691–730. https://doi.org/10.1007/s11097-018-9597-4

[33] Kathy C. Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications. London.

[34] Uwe Flick. 2018. *An Introduction to Qualitative Research* (6. izdaja). SAGE. Los Angeles, LA.

[35] Katsumi Watanabe. 2013. Teaching as a Dynamic Phenomenon with Interpersonal Interactions. *Mind, Brain, and Education* 7, 2, (Junij, 2013), 91-100. https://doi.org/10.1111/mbe.12011

[36] Kazuo Yano. 2013. The Science of Human Interaction and Teaching. *Mind, Brain, and Education* 7, 1, (Marec, 2013), 19–29. https://doi.org/10.1111/mbe.12003

[37] Gregor Geršak in Janko Drnovšek. 2016. Sensewear body monitor in psychophysiological measurements. *IFMBE Proceedings* 57, (Januar, 2016), 437–441. https://doi.org/10.1007/978-3-319-32703-7_85

[38] Jennifer A. Healey in Rosalind W. Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6, 2, (Junij, 2005), 156–166. https://doi.org/10.1109/TITS.2005.848368

# The ONE-ness of change: An exploratory neurophenomenological single case study on change in mood

Tine Kolenik[†,*]
Department of Intelligent Systems
Jožef Stefan Institute
Ljubljana, Slovenia
tine.kolenik@ijs.si

Jaya Caporusso[*]
MEi:CogSci
University of Vienna
Vienna, Austria
jaya.caporusso96@gmail.com

## ABSTRACT

The process of change is universally referred to when explaining the human psyche in the domain of attitude and behavior change. However, change is either presumed to simply exist without further elaboration, or it is reduced to neurobiological processes. While there is a substantial effort to detect, forecast and induce change, especially in the mental health-related fields, the results have been mixed so far. Understanding what change is is therefore crucial. Data on first-person experience has been thus far absent from studying change, which may turn out to be a deciding oversight. This exploratory study employs the framework of neurophenomenology to explore the process of change from multiple perspectives. In this circularly informing research process, we used ecological momentary assessment to gather daily questionnaire and diary data on mood. Afterward, we selected a single case, and determined the moment of change in mood through an inter-methodological agreement using qualitative and computational methods. Lastly, we conducted phenomenological interviews to study change on the experiential level. We found that while there may be inter-methodological agreement on the moment of change, different levels of analysis (operational, narrative, experiential - ONE) establish different definitional aspects, whereas the existence of change on the experiential level is unclear. It was ambiguous whether the same phenomenon was studied even after inter-methodological agreement. Further intersubjective research is needed to explore the phenomenon further.

## KEYWORDS

ecological momentary assessment, empirical phenomenology, human change processes, idiographic computational dynamics, mental health, natural language processing, neurophenomenology



*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*


* The authors contributed equally to this work.

[1] Please note that throughout the text we use "state A" and "state B" to delineate, respectively, the state before change and the state after change. However, the different instances of "state A" and "state B" do not necessarily correspond.

## 1 INTRODUCTION

Established research on the mind related to human change processes, also referred to as attitude and behavior change, presumes change simply exists, without any further elaboration. Implicitly, researchers treat change as dark matter: there is *state of interest* $S_A$ at *a time t*, *state of interest* $S_B$ at *time t+1*, and what happens in between is magic [1-4].[1] When change is defined, albeit rarely, it falls into reductionist pits, being reduced to neurobiological processes [5], or it is defined functionally, where change equals $S_B$ less $S_A$, especially in quantitative research.

Thus, research is mostly concerned with how to drive $S_A$ to $S_B$, tackling questions such as "What motivates change?", "How is change implemented?", "How is change sustained?", "When to induce change?", and similar [1-4]. What surprisingly lacks from this list is a bit more intimate and primary: **What is change?**

The question is neither trivial nor unimportant. Various domains interested in change - from mental health [4] to green behavior [6] - are facing a considerable obstacle when trying to detect, forecast and induce (desired) change [7]. Physiological (e.g., sensors) and psychological (e.g., questionnaires) tools have been used to this end, but have produced mixed results, especially on longer scales [8]. What is more, it seems that cognitive science is still in its infancy when studying change. Analogies can be found in both extreme levels of analysis. In physics, classical thermodynamics ignored the process of change, and it was only non-equilibrium thermodynamics that started to consider change as a fundamental process as opposed to only studying substances [9]. In philosophy, process philosophy faced Plato's claim on change as illusionary, and stood against the classical philosophical view of ontology [10]. Post-cognitivist paradigms in cognitive science provided a similar opposure, especially dynamical systems theory (e.g., psychotherapy [7]). In behavioral sciences, the study of persuasion is starting to brush against the notion of what change might be [11].

Another consequence of the prolificacy of post-cognitivist paradigms was the introduction of first-person experience [12] as an essential aspect of studying the mind. Expectedly, empirical phenomenology [13, 14] has so far eluded inclusion into the science of change, an oversight which might have hurt its efforts.

First-person experience reports might uncover experiential patterns that may thus prove to be an invaluable tool for answering the questions on change. Phenomenological interviews in particular are often focused on the "transitions between different phases [in time of an experience]" [15, p. 6], and could therefore elucidate the nature of the *magic* happening between two states. However, to our knowledge, no empirical phenomenological study investigated the experience of change, that is, no empirical phenomenological study aimed at an accurate phenomenological description of how it is to experience change per se (for a study on the experiential nature of the transition between two sequential moments, see [16]).

This exploratory study therefore aimed to spur non-reductionist research on the fundamental nature of what change is (see section Outline of the research framework for details). The general domain of mental health offers an appropriate context to study change, because it makes it salient. We focus on change in mood, which is not only ubiquitous, but also one of the primary concerns in mental health. We followed neurophenomenology [17] on combining first-person and third-person methodologies with mutual constraints, and used ecological momentary assessment (EMA) to collect daily quantitative and qualitative data on mood as well as conducted phenomenological interviews on selected data.

## 2 OUTLINE OF THE RESEARCH FRAMEWORK

The highly exploratory nature of this research is two-fold: 1) its object of inquiry is on the one hand ubiquitous and on the other hand definitionally very vacuous; and 2) the mutual-informing of the methods used has been untested so far. Since our presupposition is that change is fundamentally a dynamical process, we rely on collecting time series and diachronic data. Due to the human idiography [18], this touches the framework of small or personalized data [19], where inter-human variance and noise are reinterpreted and feature as important data. Following this, our framework investigates a moment in time with dynamics-sensitive methods on various levels of analysis. What is sought is inter-methodological agreement, and descriptions of the phenomenon on various levels of analysis. Once the latter are gathered, the unified definitional outlines can occur.

For this study, we are focusing on a single case, and within this single case, on a single identified unity of data. We believe methodological pluralism is necessary to explore this phenomenon. Note that this research is not executed sequentially, as various types of data inform one another and the direction of the research [17]. The decision on the context of mood was made due to the ubiquity of it, and the importance of change processes in mental health. We note that change may not be invariant in every context.

### 2.1 Research Questions

This work pursues the following research questions:

**RQ1**: What is the inter-level agreement between various methods with which change can be detected?

**RQ2**: What are the properties of change that are discerned (or constructed) by various methods and where do they diverge? Are they addressing and describing the same phenomenon?

RQ1 is concerned with the level of methodological agreement that change occurred in a selected moment in time. RQ2 is concerned with how change can be described when using specific methods, how the latter influence the definition, and whether the phenomenon they ultimately research is the same.

The research questions specific to the phenomenological investigation were informed by time series data.

**pRQ1**: Was change experienced at any point of the investigated episode?

**pRQ2**: What is the experiential difference between the state before and the state after the change?

These RQs cannot be addressed through the results only due to the exploratory nature of the work. We thus partially address them in the Discussion section as well.

## 3 METHODOLOGY

To pursue the research questions, we employed a mixed-methods methodology, using quantitative data collected from daily questionnaires, text data collected from daily diary entries, and first-person experiential data collected with phenomenological interviews. Due to the circular informing that occurred between these data that guided the research, we have adopted the framework of neurophenomenology, where "'neuro' refers [...] to the entire array of scientific correlates which are relevant to cognitive science" [17, p. 330].

To be able to study change ecologically, occurring in the wild as much as possible, we followed the EMA framework, which involves "repeated sampling of subjects' current behaviors and experiences in real time, in subjects' natural environments", which aims to "minimize recall bias, maximize ecological validity, and allow study of microprocesses that influence behavior in real-world contexts" [20, p. 1].

### 3.1 Materials

We used the 10-item international Positive and Negative Affect Schedule Short Form (I-PANAS-SF) in English [21] to collect daily mood data. I-PANAS-SF evaluates the following moods in a desired time span (in our case, daily) on a 5-point Likert scale: Afraid (AF), Alert (AL), Determined (DE), Distressed (DI), Enthusiastic (EN), Excited (EX), Inspired (IN), Nervous (NE), Scared (SC), Upset (UP).

To collect the diary entry data, guidelines suggested to the co-researchers (see [22] for the use of the term co-researcher) to focus on the descriptions of mood, the effects of mood on the experiences of themselves and the world, the change of the latter from the previous day to the present day, and on any salient factual information about the day (for more, see Supplementary materials, section Diary entry guidelines).

## 3.2 Sample and Case

The sample included seven people, largely acquaintances of the authors, from which a single person was arbitrarily selected, codenamed as Quentin. Our co-researcher was 30 years old at the end of the data collection phase, biologically assigned at birth as male and identifying as a man and as non-binary, with a master's degree. He was of somewhat good mental health, had never been diagnosed with a mental disorder, did not have mental health-related therapy in the recent past, and was not taking any mental health-related medications. He slept seven hours on average per night and had bad sleep quality. He was generally a positive person who felt neutral about his emotional arousal or did not identify with having positive or negative emotional arousal. His experience with phenomenological reporting amounted to around 70 hours.

## 3.3 Data Collection

We used the Synergetic Navigation System (SNS), a web- and mobile-based technology for EMA [23], to collect questionnaire and diary data, and conducted in-depth phenomenological interviews based on micro-phenomenology [24] to collect experiential data. The data was collected from June 24th to July 14th 2021. The SNS data on a given day was collected from 18:00 onwards on the same day or in the morning of the following day. Quentin was notified at the starting hour of data collection through email and mobile push notifications. The interviews were recorded with a Samsung Galaxy A41.

## 3.4 Computational Definition of Change

To detect change in quantitative data, change had to first be defined methodologically. Since quantitative data are generally analyzed computationally, we present a computational definition of change which was applied to the data. We computationally defined (inspired from sudden gains literature [11] and anomaly detection [25]) that change *C* between data point or state A ($S_A$) at time *t* and data point or state B ($S_B$) at time *t+1* occurs if

$$(((SB > (M + SD/MAD)) \parallel (SB < (M - SD/MAD))) \;\&\; ((M - SD/MAD) < SA < (M + SD/MAD))) \quad (1)$$

$$\parallel$$

$$(((SA > (M + SD/MAD)) \parallel (SA < (M - SD/MAD))) \;\&\; ((M - SD/MAD) < SB < (M + SD/MAD))) \quad (2)$$

$$\parallel$$

$$(\Delta C > SD/MAD) \quad (3)$$

where *M* equals the mean value of the entire time series and standard deviation (*SD*) is used if the data is normally distributed, and median and mean average deviation (*MAD*) is used if the data is not normally distributed. It denotes that change occurs:

a) if $S_B$ falls outside of bounds of *SD*/*MAD* while $S_A$ falls inside (Equation (1)); or

b) if $S_A$ falls outside of bounds of *SD*/*MAD* while $S_B$ falls inside (Equation (2)); or

c) if both are inside the *SD*/*MAD* bounds, $S_B$ is more than one *SD*/*MAD* away from $S_A$ (Equation (3)).

*SD*/*MAD* bounds represent the baseline, which means that between a data point falling outside of these bounds while the preceding data point was inside the bounds change occurred, and vice versa. If inside the baseline, change can still occur, but it has to be bigger than one *SD*/*MAD*.

To apply this calculation to the data, it has to be preprocessed, extracting the described values.

This computational definition of change is independent of the context (in our case, mood).

## 3.5 Empirical Phenomenology

We included empirical phenomenology as a method to obtain data on experience. Empirical phenomenology, based on the concept of *epoché* [26], allows to get descriptions of how the investigated episodes and phenomena are actually lived. It excludes the possible narratives, conceptualizations, and judgements that might be constructed after the experience per se. In particular, we opted for an interviewing approach based on the micro-phenomenological interview method [24]. The interviewer non-suggestively accompanies the interviewee in providing accurate phenomenological descriptions of the diachronic (temporal unfolding) and synchronic (non-temporal dimension, associated with a specific moment or phase) structure of the experience. For these reasons, empirical phenomenology allowed us to investigate how it is to experience change. The interviewing was informed by our research questions, and the interviews were conducted after change had already been partially identified (see Results, Subsection Identifying the moment of change).

## 3.6 Collected Data

Quentin completed 16 questionnaires and provided 16 diary entries between June 24th and July 14th 2021. The mean of Quentin's diary entries was 195 words. Furthermore, three in-depth phenomenological interviews were conducted on the selected moment within the time series data (see Results, subsection Identifying the moment of change), clocking 00:43:33, 01:00:51, 1:09:41 in length, respectively. The interviews are being transcribed verbatim.

# 4 RESULTS

This section presents the analysis of the collected data. For computational analysis of the time series data, features were extracted in order to calculate whether and when change had occurred. Change was already defined computationally for the time series in the previous section. Phenomenological results present the effort to identify change experientially, without a definition pre-given by the authors.

## 4.1 Feature Extraction

Features extracted from the quantitative questionnaire scores included: the mean and SD of a given mood category if the distribution was normal; and the median and MAD if the distribution was non-normal. We performed a normalcy test to discern that. Instances with missing questionnaire data were filled with last collected data scores.

Features extracted from the diaries included sentiment analysis features and statistical features of sentiment features. We used VADER, "a lexicon and rule-based sentiment analysis tool," [27, para. 1] to get negative and positive sentiment scores for each daily diary entry. Afterward, we extracted statistical features following the same process as for the quantitative questionnaire scores. Instances with missing diary data and therefore missing sentiment scores were linearly interpolated.

## 4.2 Identifying the Moment of Change

To identify the moment of change and address RQ1, several steps were taken. Authors studied the data, particularly reading the diary entries, and asked Quentin to propose a data instance where he felt an instance of change had occurred. Quentin suggested the data instance from July 1st, 2021. This is the selected data instance:

**Table 1: Quentin's mood scores on July 1st 2021 (see full names in this subsection, para. 5).**

| DI | AF | UP | NE | SC | IN | AL | EX | EN | DE |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 5 | 5 | 5 | 5 | 5 |

For the selected diary entry (DiaryE0) and the data from the previous day, see Supplementary material, section Diary entries and quantitative questionnaire mood scores. The text part (pDiaryE0) containing the description of change can be read below:

> *I saw myself as important, I was very self-confident. This brought about a certain feeling, a certain change in the air around me. [...] people listening to me [...] had this directionality towards me which gave me some sort of power. Compared to yesterday, when I also felt inspired and enthusiastic, today I had this huge undertone of confidence, and this caused a difference especially in how I perceived others.*

Quentin confirmed this is a good example of change occurring during the data collection. The authors had beforehand identified the same data instance as a potentially good candidate. The change specifically referred to the particular confidence ("*Compared to yesterday, […] today I had this huge undertone of confidence*"). The state of the mood before the change (or State A) was therefore either no confidence or a different kind of confidence, coupled with inspiration and enthusiasm, and the state of the mood after the change was the newly found confidence (or State B).

For inter-methodological agreement on the moment of change, the computational method for detecting change (see Methodology, subsection Computational definition of change) was applied to two data streams, the quantitative questionnaire scores (all the 10 mood categories) and the diary entries.

For the quantitative questionnaire scores, change was detected in 7 out of 10 mood categories (AL, DE, DI, EN, EX, IN, UP). The three categories where change was not detected (AF, NE, SC) were stationary, which means that there were no changing curves. See Supplementary material, section Mood graphs for mood graphs with detected change. Figure 1 presents one such graph, signifying the change in EN.


**Figure 1: Detected change in the Enthusiastic mood category from 30. 06. 2021 to 01. 07. 2021.**

Furthermore, change was detected in both the positive and negative sentiment scores from the diary entries. See Supplementary material, section Sentiment graphs for sentiment graphs.

The results show maximum inter-methodological agreement. Every part of the two data streams that could have possibly validated the initial identification of change had validated it. The next step was to see whether change occurred in the selected moment on the experiential level.

## 4.3 Phenomenological Results

To identify the moment of change to be investigated in the interviews, we analyzed the fragment of the co-researcher's diary entry in which the selected episode is described (pDiaryE0 below Table 1). We identified two possible instances of change: one in the third sentence, and the other in the last sentence. We decided to focus on the first one, as it seemed to have had occurred at a specific point in time, and it was therefore possible to investigate it with phenomenological interviews.

We present the provisional results of the phenomenological investigation. On the 1st July 2021, our co-researcher, Quentin, was giving a lecture at a seminar. He was sitting at a desk in a lecture hall, and he was talking to the people in front of him. He initially felt a self-confined confidence that later changed into a new confidence. We summarize the experiential categories that were different from before (state A) to after the change (state B) in Figure 2 (pRQ2).

In between state A and state B, Quentin noticed a ray of sunshine filtering through the air. He felt like his arms had the potentiality to move more freely in that direction, experienced as a sense of brightness on their upper left part. This aspect was part of the new confidence, which was not fully present yet. Quentin realized that this brightness was something new ((1) in Figure 2). Quentin felt a ball-like entity in his chest, which expanded until it reached the audience. It is at this point that the experience reached state B, where Quentin felt the full new confidence. Quentin had the knowledge that the way he was perceiving and could interact with people had changed ((2) in Figure 2).

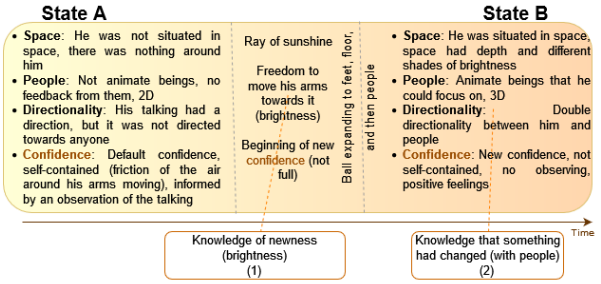



**Figure 2: Experiential structure (diachronic and synchronic) of the target episode.**

# 5 DISCUSSION

## 5.1 The ONE-ness of Change

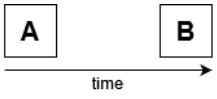
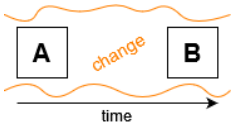
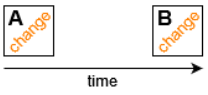
While discussing the data from the different methods, we specified incongruencies between the data and how it characterizes change (addressing RQ2). In quantitative analysis, change was necessarily defined by the authors - the computationally defined bounds were arbitrary wrt the phenomenon itself. We labeled this kind of (definition of) change OPERATIONAL (definition of) change (oC). In diary data, change was defined by the co-researcher in two instances, one of which includes the exact word "change" (see pDiaryE0). It is argued that we "organize [our] experiences and actions according to narrative structures thereby situating them in the context of a unifying story," [28, p. 179] which we attest also happens while writing a diary entry. Change was therefore narratively constructed. Arguably, this construction occurred in the moment of the writing of the diary, at a point in time successive to the original experience that the narrative was about. We labeled this kind of change NARRATIVE change (nC). In the phenomenological data, change was looked for in a collaboration between the co-researcher and the researcher conducting the interviews. Differently from the other levels of analysis, the understanding of something as change was here not already given, but to be explored and discussed. In fact, our phenomenological inquiry was aimed precisely at investigating how change might present itself in experience, if it does at all. We labeled this kind of change EXPERIENTIAL change (eC).

There are two big problems that arise from this: a) the granularity problem, and b) the level problem. It is not clear how the various time spans correspond to each other (a)), and whether the various levels of change (oC, nC, eC) refer to and describe the same phenomenon, using different levels of analysis. It might in fact be that oC, nC and eC refer to multiple phenomena. We do sense there is a certain correspondence between the three levels (and there was an agreement on the moment between the co-researcher's suggestion and the authors' suggestion), but unraveling the complexity of that is out of scope of this paper.

## 5.2 Models of Experiential Change

We hypothesize different models of how change might be experienced in a simplified "state A to state B transition".

**Table 2: Models of experiential change.**

| | |
|---|---|
| 1) Change is not present at all at the level of experience. |   |
| 2) Change is the experiential nature of the experiential flow in which state A and state B succeed each other. |   |
| 3) Between state A and state B there is a state C where change is experienced. |   |
| 4) a) Change is an experiential element present in both state A and state B. b) Change is an experiential element present either in state A or c) in state B. |   |

Some representational aspects of the models above are due to functional reasons. We envisioned further models but for the sake of brevity we only included some.

Following, we discuss how we tried to address pRQ1. During the first interview, Quentin said: "*Not that I felt the change, the change happened and I felt the consequences of the change*". This seems to suggest either model 1) or 4c). Later, we found two different instances of experience that could represent experiential change. The first refers to (1) in Figure 2. Quentin made it clear that the knowledge was about the brightness being something new, not something different from before, since "*There was no trace of what was before or how this came to be*". This does not mean that this experience does not entail experiential change: as far as we know, experiential change might be precisely experienced as the knowledge, or perception, of the newness of something. This would correspond to model 3). The other instance that might delineate experiential change refers to (2) in Figure 2. This change would correspond to model 4c). However, we were specifically interested in the experience of change in mood, and we cannot claim that the change referred to in (2) in Figure 2 complies with this. When asked towards the end of the third interview whether at any point of the investigated episode he realized that his confidence had changed, Quentin answered no (which hints at model 1)).

# 6 CONCLUSIONS AND FUTURE WORK

This work represents an exploratory neurophenomenological inquiry into the nature of change in the context of mood. We used ecological momentary assessment to collect daily questionnaire and diary data, and after selecting a proper data instance, we conducted phenomenological interviews on it. We discerned that there was an inter-methodological agreement on the moment of change; however, it is not clear how it manifested, if at all, on the experiential level. We observed various definitional aspects of change, culminating in *ONE-ness of change*, describing operational, narrative, and experiential change. Finally, we presented some possible models of experiential change and

analyzed how our phenomenological data fit into them. We found two major problems to address in the future: the granularity and the levels problem.

The study had many limitations, mostly due to its exploratory nature. It was ultimately single case, where it analyzed only one episode. It had a limited number of interviews, which may have not gone in depth enough to really identify and specify the phenomenon of interest. Furthermore, interviews on the moment of writing might be necessary as well. When collecting quantitative data, not every day was sampled, and the amount of data may have produced biased baseline calculations, resulting in faulty change detections. Using a single method to detect change may also not be enough, and a discussion is needed on how to proceed when two methods from the same or different levels of analysis disagree on the change moment. We will not delve into the potential problems of ecological momentary assessment and quantitative and qualitative self-reports.

In future work, apart from addressing the limitations, we plan to continue with the general effort of this study. Future possibilities include: applying the same methodology transdiagnostically and for induced, volitional and spontaneous change; conducting interviews on episodes reported as including experiential change, and with expert meditators observing change; analyzing the inter-methodological and experiential structure of change, where it seems to follow some aspects of the matryoshka principle [29]; applying post-cognitivist frameworks, e.g., the dynamical systems theory framework; addressing the granularity problem by expanding the methodology by changing the EMA contingency (e.g., when experiential change occurs, when a physiological signal occurs) and including descriptive experience sampling [30]; seeing whether change can be forecasted with machine learning and what implications it brings; exploring what the possibilities in how oC, nC and eC relate to one another are; testing models of experiential change with computational simulations; making the dataset and codebook publicly available; interpreting our findings in the contexts of different theories of change and time.

## ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209; the Young researchers' grant), and thank their wonderful co-researchers.

## REFERENCES

[1] Charles Abraham and Susan Michie, 2008. A taxonomy of behavior change techniques used in interventions. *Health Psychology*, 27, 3 (May, 2008), 379–387. DOI: https://doi.org/10.1037/0278-6133.27.3.379

[2] Robert Cialdini, 2017. *Pre-suasion*. Random House Business, New York, NY.

[3] Brian J. Fogg. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann, Burlington, MA.

[4] Richard H. Thaler and Cass R. Sunstein, 2021. *Nudge: Improving decisions about health, wealth and happiness*. Yale University Press, New Haven, Connecticut.

[5] John T. Cacioppo, Stephanie Cacioppo, and Richard E. Petty, 2018. The neuroscience of persuasion: A review with an emphasis on issues and opportunities. *Social neuroscience*, 13, 2 (Mar 2018), 129-172.

[6] Fredrik Carlsson, Christina Gravert, Olof Johansson-Stenman, and Verena Kurz, 2021. The Use of Green Nudges as an Environmental Policy Instrument. *Review of Environmental Economics and Policy*, 15, 2 (Jun 2021), 216-237.

[7] Günter K. Schiepek, Kathrin Viol, Wolfgang Aichhorn, Marc-Thorsten Hütt, Katharina Sungler, David Pincus and Helmut J. Schöller, 2017. Psychotherapy is chaotic — (not only) in a computational world. *Frontiers in Psychology*, 8 (May 2017), 379.

[8] Tine Kolenik, 2021. Methods in Digital Mental Health: Smartphone-based Assessment and Intervention for Stress, Anxiety and Depression. In *Integrating Artificial Intelligence and IoT for Advanced Health Informatics*. Springer, Berlin. In print.

[9] Howard Reiss, 1996. *Methods of thermodynamics*. Courier Corporation, Chelmsford, MA.

[10] Nicholas Rescher, 1996. Process metaphysics: An introduction to process philosophy. Suny Press, Albany, NY.

[11] Tony Z. Tang and Robert J. DeRubeis, 1999. Sudden gains and critical sessions in cognitive- behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, 67, 6 (Dec 1999), 894–904. DOI: https://doi.org/10.1037/0022-006X.67.6.894

[12] Joana Rigato, Scott M. Rennie, and Zachary F. Mainen, 2019. The overlooked ubiquity of first-person experience in the cognitive sciences. Synthese, (Feb 2019), 1-37.

[13] Patrik Aspers, 2009. Empirical phenomenology: A qualitative research approach (The Cologne Seminars). *Indo-pacific journal of phenomenology*, 9, 2 (Oct 2009), 1-12.

[14] Urban Kordeš, 2016. Going beyond theory. *Constructivist Foundations*, 11, 2 (Mar 2016), 375-385.

[15] Jaya Caporusso and Ema Demšar, 2020. Phenomenology of dissolution experiences: An exploratory study. In *Proceedings of 23rd International Multiconference INFORMATION SOCIETY (Vol. B).* T. Strle, J. Černe, & O. Markič (Eds.), Institut "Jožef Stefan", Ljubljana, Slovenia, 5-9. https://www.academia.edu/44494880/Phenomenology_of_Dissolution_Experiences_An_Exploratory_Study_conference_contribution_

[16] Julian Bass-Krueger, 2021. *Consciousness and Time* [Unpublished master's thesis]. University of Vienna.

[17] Francisco J. Varela, 1996. Neurophenomenology: A methodological remedy for the hard problem. *Journal of consciousness studies*, 3, 4 (Apr 1996), 330-349.

[18] Peter CM. Molenaar, 2004. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement, 2, 4 (Oct 2004), 201–218. DOI: https://doi.org/10.1207/s15366359mea0204_1

[19] Rutvik V. Shah, Gillian Grennan, Mariam Zafar-Khan, Fahad Alim, Sujit Dey, Dhakshin Ramanathan, and Jyoti Mishr, 2021. Personalized machine learning of depressed mood using wearables. *Translational Psychiatry*, 11, 1 (Jun 2021). DOI: https://doi.org/10.1038/s41398-021-01445-0

[20] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford, 2008. Ecological momentary assessment. *Annu. Rev. Clin. Psychol*., 4 (Apr 2008), 1-32.

[21] Edmund R. Thompson, 2007. Development and validation of an Internationally Reliable short-form of the positive and negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38, 2 (Mar 2007), 227–242. DOI: https://doi.org/10.1177/0022022106297301

[22] Urban Kordeš and Florian Klauser, 2016. Second-person in-depth phenomenological inquiry as an approach for studying enaction of beliefs. *Interdisciplinary Description of Complex Systems: INDECS*, 14, 4 (Oct 2016), 369-377.

[23] Günter Schiepek, Heiko Eckert, Benjamin Aas, Sebastian Wallot, and Anna Wallot, 2015. *Integrative psychotherapy: A feedback-driven dynamic systems approach*. Hogrefe Publishing, Göttingen. DOI: https://doi.org/10.1027/00472-000

[24] Claire Petitmengin, 2006. Describing one's subjective experience in the second person: An interview method for the science of consciousness. Phenomenology and the Cognitive sciences, 5, 3 (Dec 2006), 229-269. DOI: https://doi.org/10.1007/s11097-006-9022-2

[25] Mark Woodard, Michael Wisely, and S. Sedigh Sarvestani, 2016. A survey of data cleansing techniques for cyber-physical critical infrastructure systems. *Advances in Computers*, 102, 63–110. DOI: https://doi.org/10.1016/bs.adcom.2016.05.002

[26] Edmund Husserl, 1983. *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy* (F. Kersten Trans.). Martinus Nijhoff Publishers, Leiden. (Original work published 1952)

[27] Clayton Hutto and Eric Gilbert, 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI.

[28] Dan Zahavi, 2007. Self and other: The limits of narrative understanding. *Royal Institute of Philosophy Supplements*, 60 (May 2007), 179-202.

[29] Wikipedia contributors, 2021. *Matryoshka doll*. In Wikipedia, The Free Encyclopedia. Retrieved 16:44, September 15, 2021, from https://en.wikipedia.org/w/index.php?title=Matryoshka_doll&oldid=1030022988

[30] Russel T. Hurlburt and Sarah A. Akhter, 2006. The Descriptive Experience Sampling method. *Phenom Cogn Sci*, 5 (Nov 2006), 271–301. DOI: https://doi.org/10.1007/s11097-006-9024-0

# Supplementary material

## 1 Diary entry guidelines

Please, answer the following questions in the form of a diary entry. Be mindful that your entry is approximately 150 words at minimum. There is no upper word limit. Questions:

1) Describe your mood.
2) Describe how your mood affected your experience:
   a) of yourself;
   b) towards the world and its elements.
3) Describe how these experiences have changed from yesterday to today.
   a) Change of experience towards yourself from yesterday to today.
   b) Change of experience towards the world and its elements from yesterday to today.
4) Factual information from the last day that you would like to highlight.

## 2 Diary entries and quantitative questionnaire mood scores

a) July 1st 2021

DiaryE0:

*Today I mostly felt quite inspired, determined and enthusiastic. I saw myself as important, I was very self-confident. This brought about a certain feeling, a certain change in the air around me. The air was pointing up, and I could move throughout differently. Also, for example, people listening to me at a seminar about using a tool for daily assessment had this directionality towards me which gave me some sort of power. Compared to yesterday, when I also felt inspired and enthusiastic, today I had this huge undertone of confidence, and this caused a difference especially in how I perceived others. Otherwise it was a full day, I had a meeting about the future of my software, I worked on my study, I sorted out the details of my stay in another country which I was invited to visit, to see the psychiatric processes and to share knowledge at their clinic, I had the before mentioned seminar, my girlfriend Jaya and I went together to a wonderful classical concert with my parents, and finally, we ate homemade apple pie and drank champagne that was a gift from my mom's best friend. It was a great day.*

Mood scores:

| DI | AF | UP | NE | SC | IN | AL | EX | EN | DE |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 5 | 5 | 5 | 5 | 5 |

b) June 30th 2021:

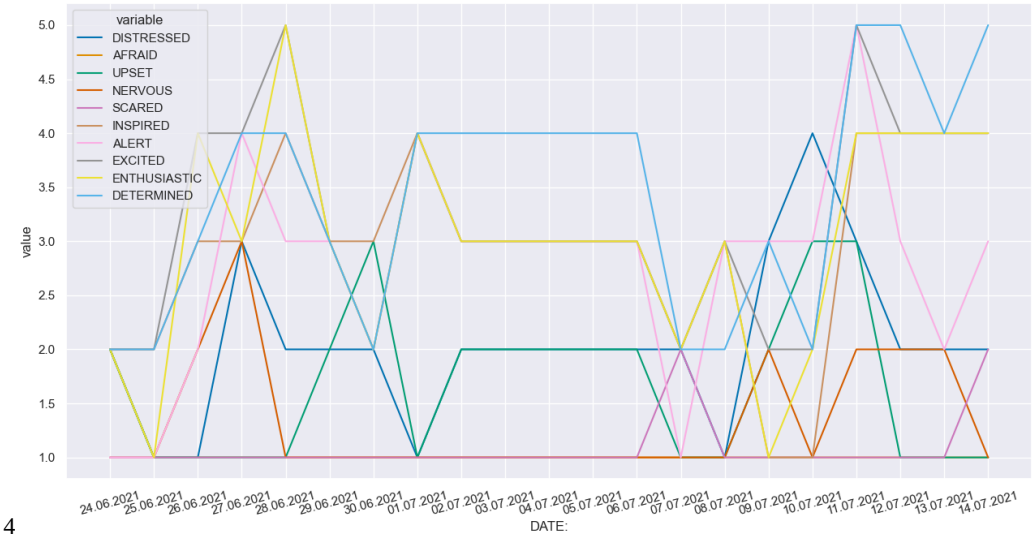
DiaryE1:

*The day was signified by three moods - uninterested, determined, and inspired. I saw myself too scattered, without a center to hold me or to hold onto, and this made me uninterested in the world around me and it was hard to do anything I wanted to, which I disliked. The narrative of where I am was quite dispersed, and it was hard to look at the things that hold me together. At first I was frustrated, so I spend some time just embracing that feeling, with knowledge that afterwards I will pull myself together. When this phase came, I became determined to set myself straight, and I made a sort of a plan or a diagram of what I want to do and is important to me, what makes me happy. This was quite successful and afterwards I was inspired to do the tasks I wanted to do. The world was consequently also different, it is like after being inspired I am seeing it, it has this brighter quality, but not visually, but the feeling of its atmosphere. Otherwise I was quite happy to have my weekly meet with two of my friends online.*

Mood scores:

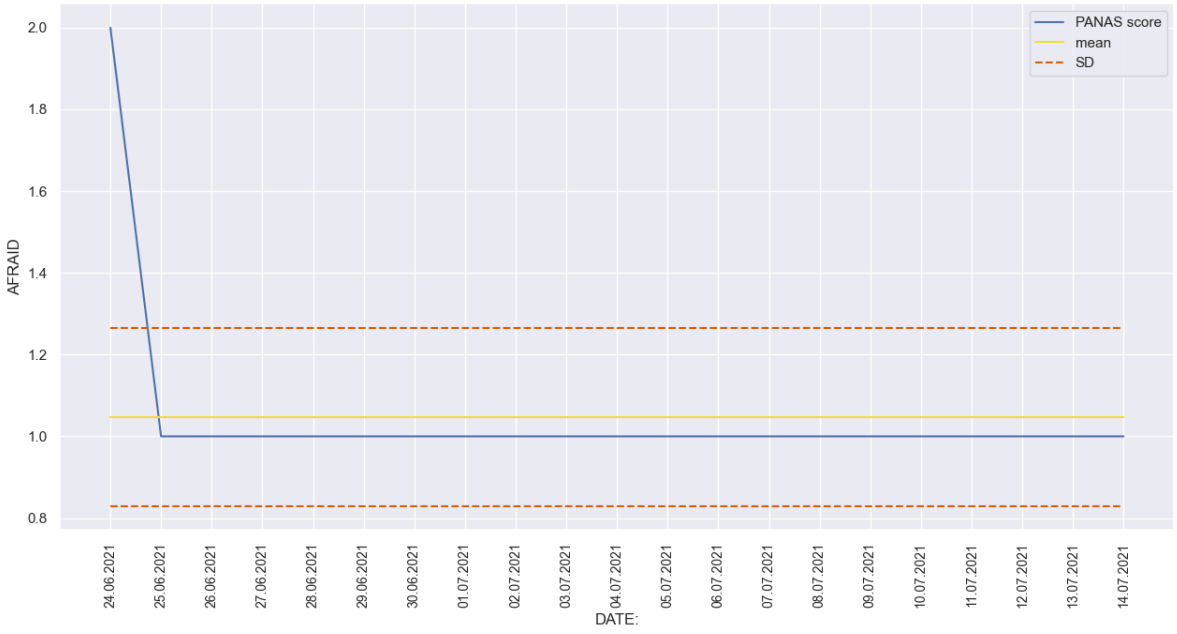
| DI | AF | UP | NE | SC | IN | AL | EX | EN | DE |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 3 | 1 | 1 | 3 | 2 | 2 | 2 | 2 |

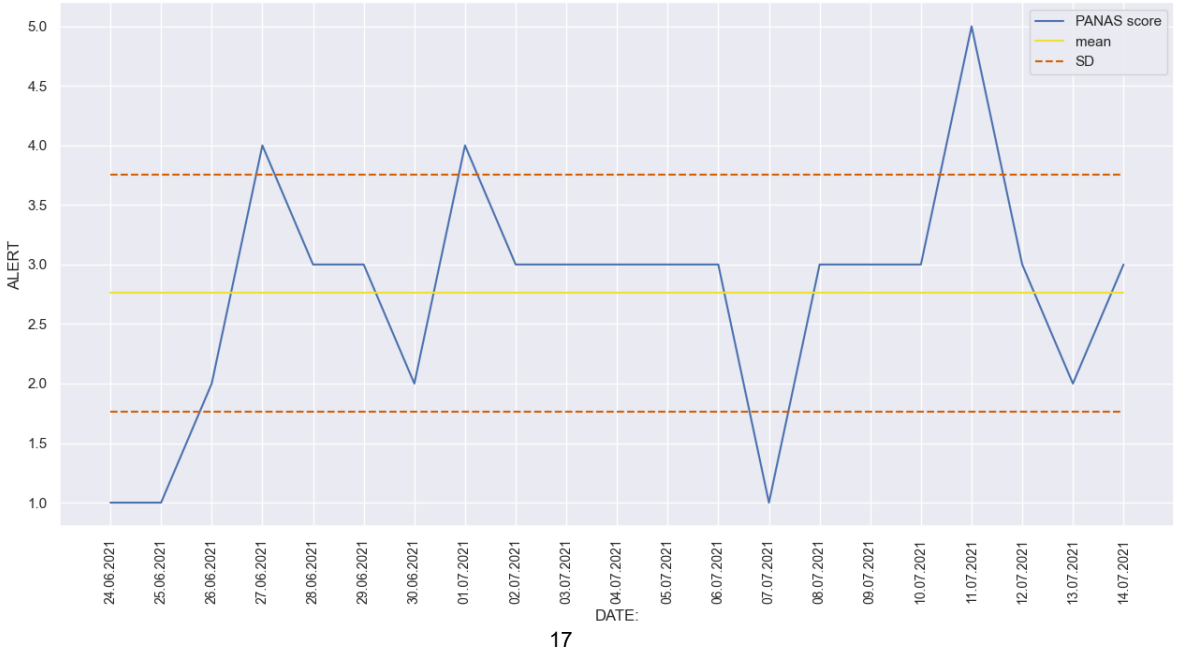
## 3 Graph of the mood over the entire time series

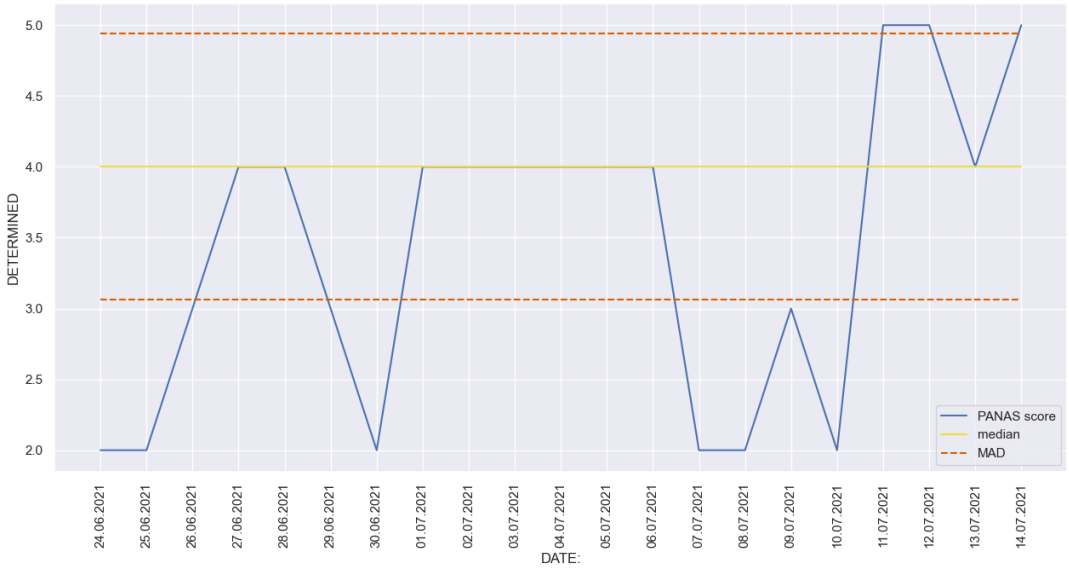


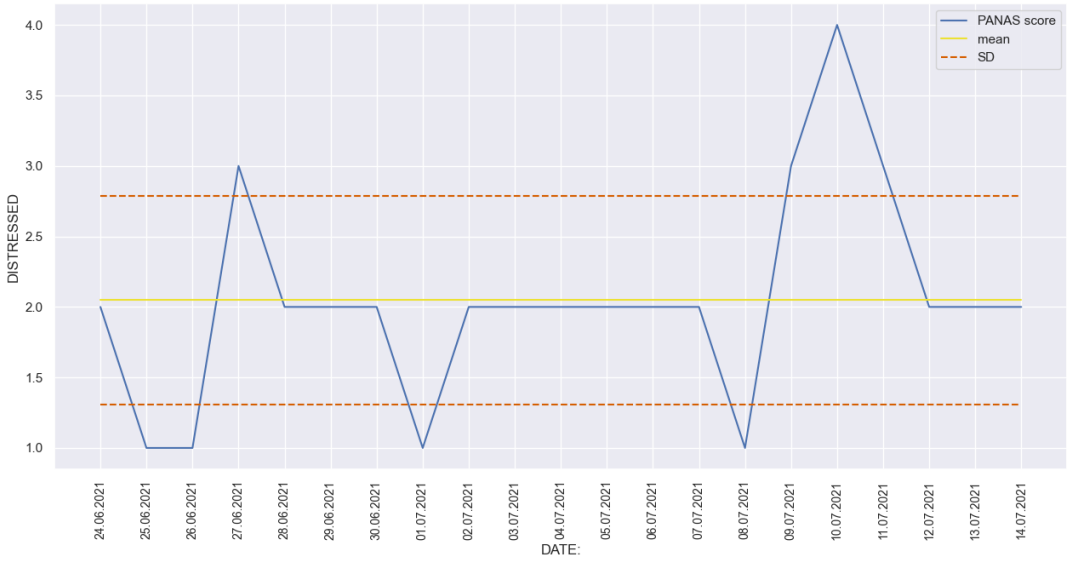

4

## 4 Mood graphs

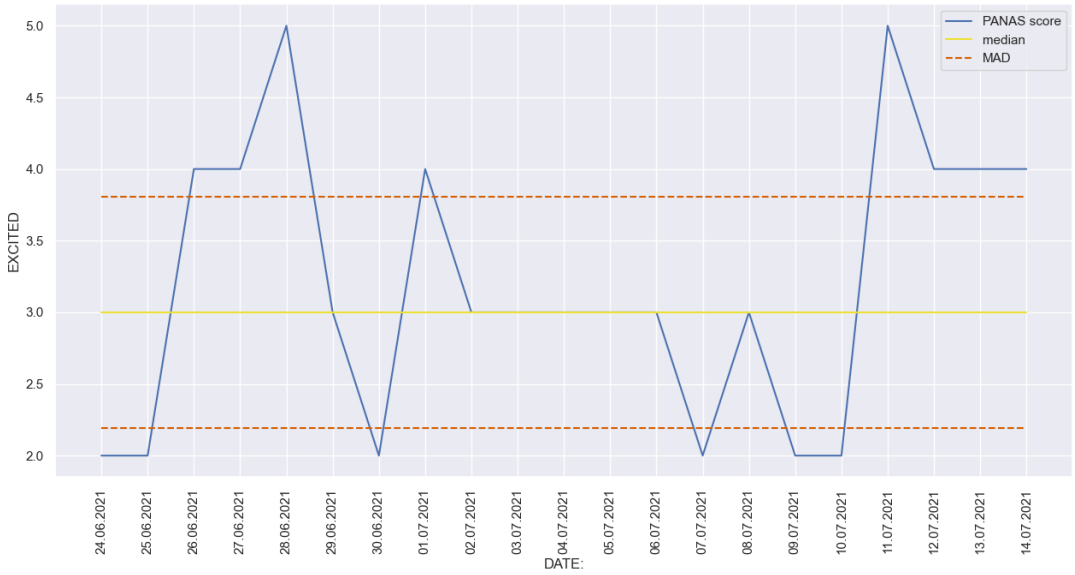


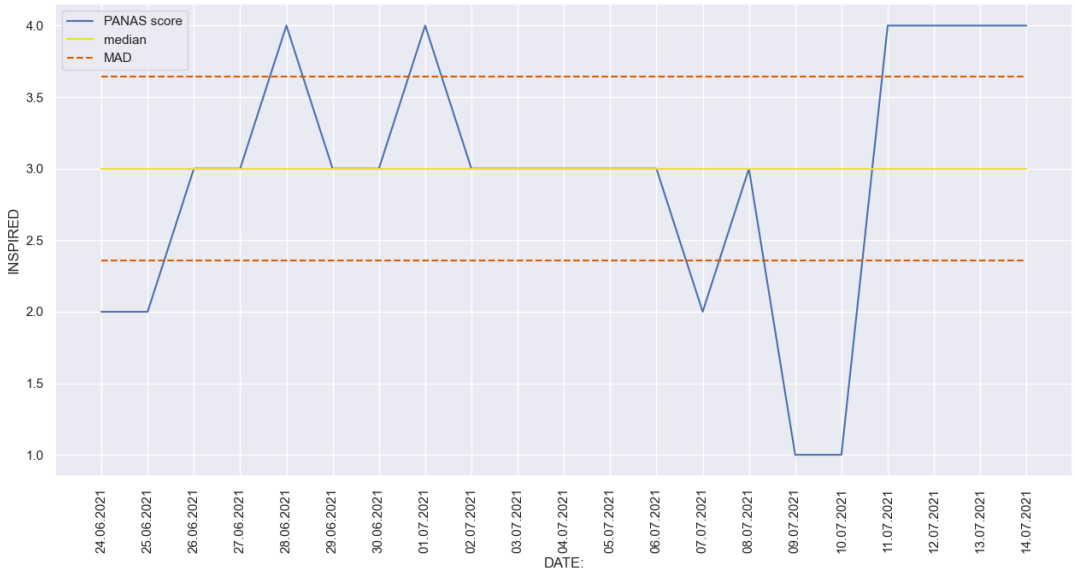



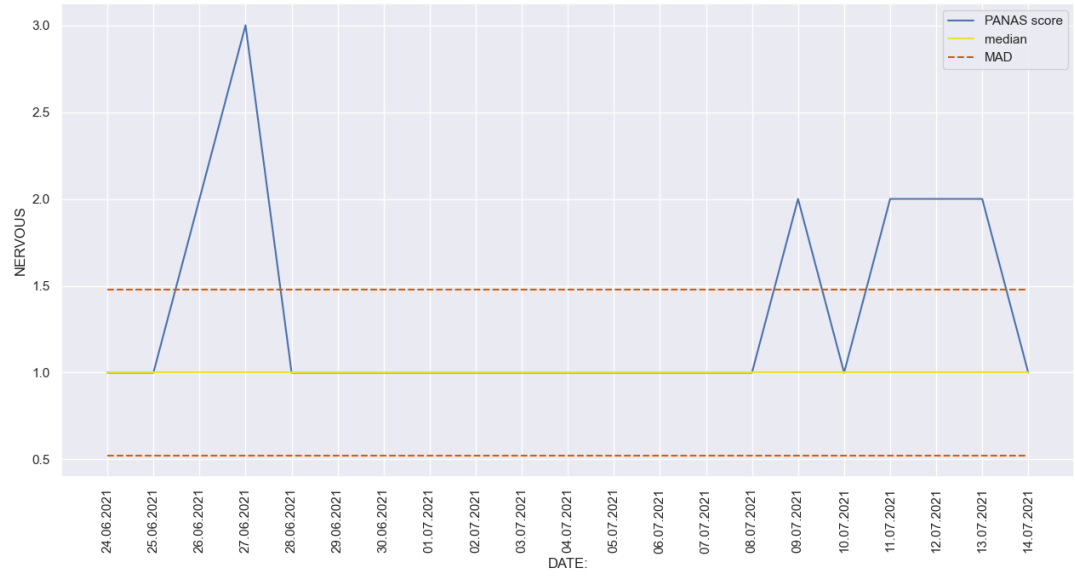

DETERMINED
5.0
4.5
4.0
3.5
3.0
2.5
2.0
PANAS score
median
MAD
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:

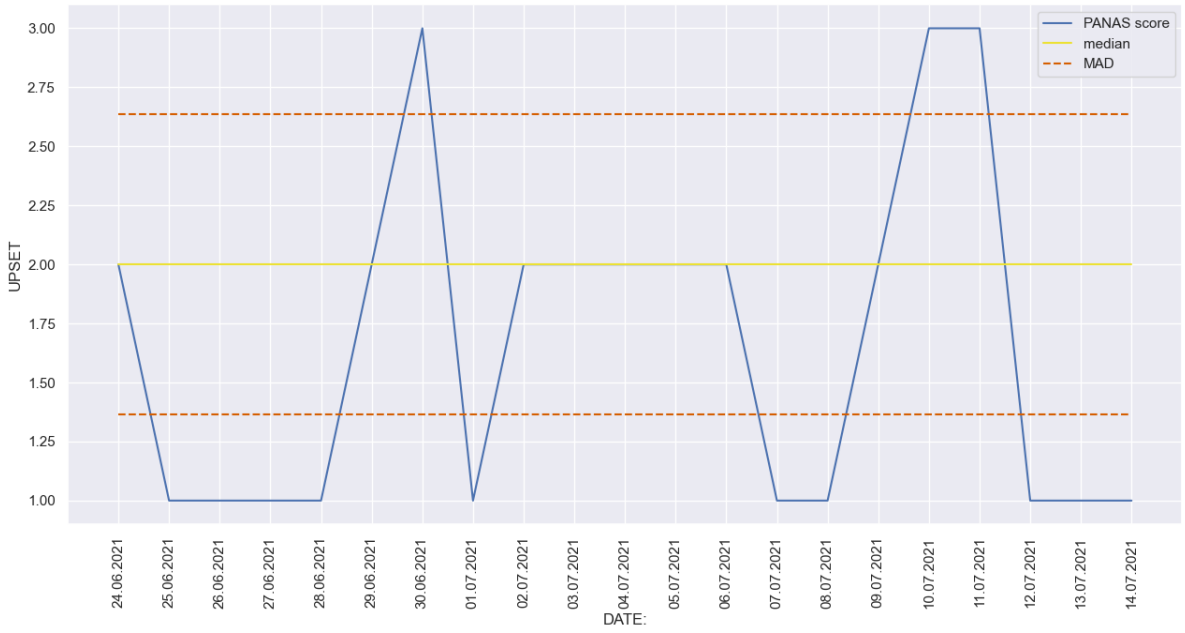

DISTRESSED
4.0
3.5
3.0
2.5
2.0
1.5
1.0
PANAS score
mean
SD
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:

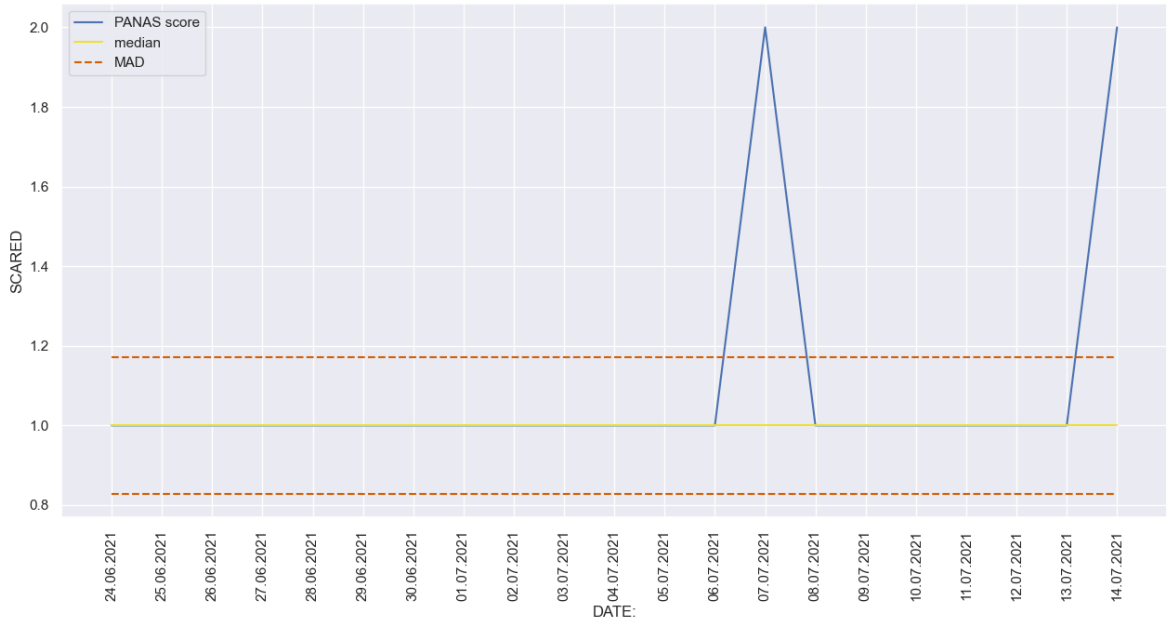

ENTHUSIASTIC
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
PANAS score
median
MAD
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:


PANAS score
median
MAD
EXCITED
5.0
4.5
4.0
3.5
3.0
2.5
2.0
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:

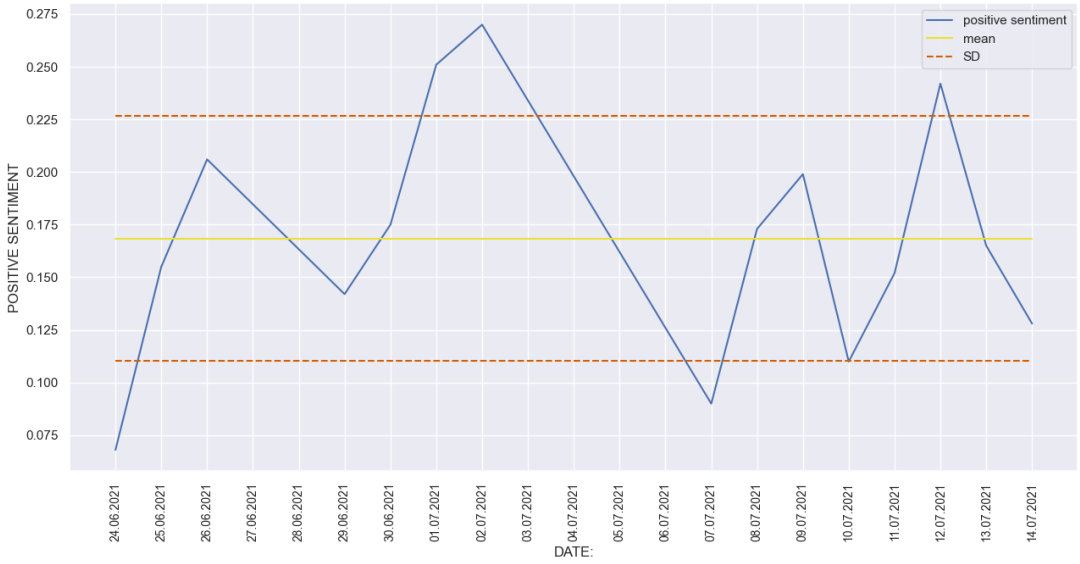

PANAS score
median
MAD
INSPIRED
4.0
3.5
3.0
2.5
2.0
1.5
1.0
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:

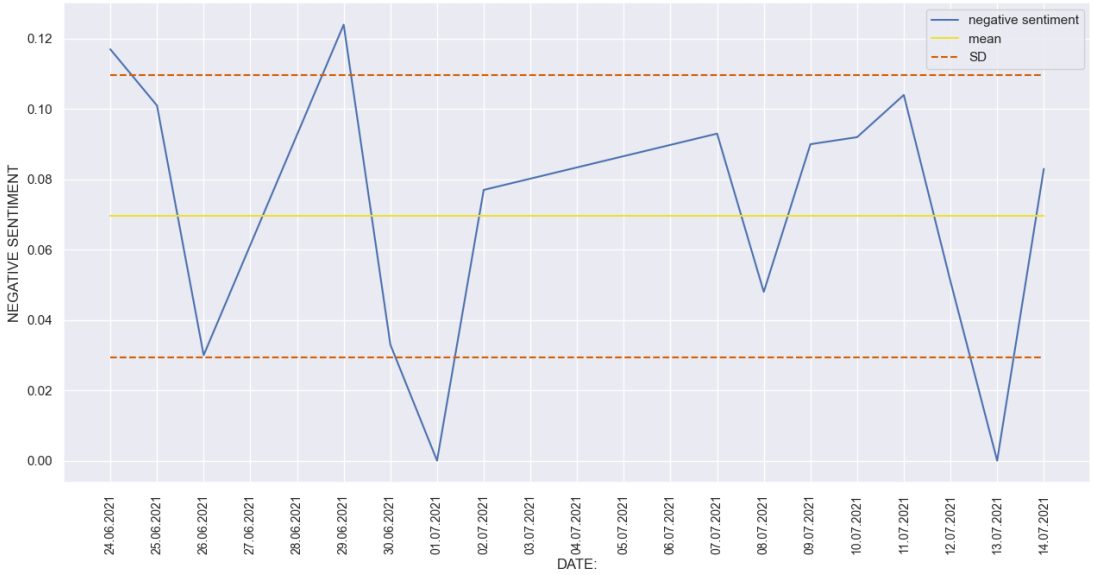

PANAS score
median
MAD
NERVOUS
3.0
2.5
2.0
1.5
1.0
0.5
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:


PANAS score
median
MAD
UPSET
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:


PANAS score
median
MAD
SCARED
2.0
1.8
1.6
1.4
1.2
1.0
0.8
24.06.2021
25.06.2021
26.06.2021
27.06.2021
28.06.2021
29.06.2021
30.06.2021
01.07.2021
02.07.2021
03.07.2021
04.07.2021
05.07.2021
06.07.2021
07.07.2021
08.07.2021
09.07.2021
10.07.2021
11.07.2021
12.07.2021
13.07.2021
14.07.2021
DATE:

# 5 Sentiment graphs



# Sensitivity of expected civilization longevity models

Anže Marinko
anze.marinko@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

Maša Žaucer
masa.zaucer@student.fmf.uni-lj.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

David Susič
David.Susic@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

Matjaž Gams
Matjaz.Gams@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

## ABSTRACT

In this paper, we analyse the parameter sensitivities of the Sandberg and Rare Earth civilization longevity models. The Sandberg model relies on the Drake equation, while the Rare Earth model assumes that the Earth is a very unique planet because of rare sequence of events causing its evolution. In addition to the sensitivity of the parameters, we also analyse the importance of those parameters.

## KEYWORDS

Human extinction, Drake equation, Civilization collapse, Rare Earth hypothesis, distributions

## 1 INTRODUCTION

After years of dealing with Fermi's question: "Where is everybody?", we still do not seem to have a good answer. After scanning more than 10 million stars [11], we have not found a single extraterrestrial life.

We know that it is inevitable that human civilization will one day die out, but what is the expected longevity and how is it related to the absence of observed civilizations? One way is to design human longevity models that use a variety of parameters to answer this question. However, it is not clear which models heavily rely on the values of parameters. In this paper we study the sensitivity of the models to the parameters and we also try to determine which parameters have the greatest impact.

In our previous papers [6, 14] we approached the topic of the extinction of human civilization and introduced the Drake equation [1]. In the first paper [6] we presented Sandberg's [8] interpretation of the Drake equation and analysed it. In the second paper [14], we presented possible causes of human extinction and used the Drake equation to estimate the longevity of human civilization. In the last paper [4], we presented four different models with some modifications of the Drake equation and considered their prospects for the time we have left. We concluded that we are most likely to survive at most 10 000 years.

In this paper, we focused mainly on two of the models from the previous paper [4]. The first model we analysed is based on Sandberg [8] and the second one represents the "rare Earth" hypothesis [12]. For both models we analysed the difference between using log-uniform and log-normal distributions of the parameters. In addition, we analysed which parameters most affect the results in each model. All in all, we dove into the structure of the models and tried to improve the accuracy of the results.


*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*


## 2 RELATED WORK

Some publications suggest there are 600 to 40 000 technological civilizations in our galaxy [10], while others think there should be about 36 of them, assuming an average lifespan of 100 years [13]. However, given our ability to detect intelligent life [3] and their radio signals [2], and the fact that we have not detected anything yet, a large number of civilizations is unlikely.

In our previous paper [4], we analyzed 4 different models of the modified Drake equation to determine longevity of human civilization. From the accessible data, we concluded that the human technological civilization will most likely survive at most 10 000 years. Note that the analysis is not able to conclude anything about biological aspects of humans. Another research induces that the yearly probability for extinction is most likely less than 1 in 87 000 using four different models [9]. In [5] they explain that humanity will eventually have to move to avoid the death of our Sun.

In this paper we focused on how the parameters of the Drake equation and the choice of the various attributes in two models affect the probability of longevity of human technological civilization.

## 3 ESTIMATING THE LONGEVITY OF HUMAN CIVILIZATION WITH SANDBERG AND RARE EARTH MODEL

### 3.1 SANDBERG MODEL

The Sandberg model [8] is based on Drake equation:

$$N = R_* f_p n_e f_l f_i f_c L \tag{1}$$

- $R_*$ being the rate of star formation per year,
- $f_p$ the fraction of stars with planets,
- $n_e$ the number of Earth-like (or otherwise habitable) planets per a star that has planets,
- $f_l$ the fraction of habitable planets with actual life,
- $f_i$ the fraction of life-bearing planets that develop intelligence,
- $f_c$ the fraction of intelligent civilizations that are detectable,
- $L$ the average longevity of such civilizations.

**Table 1: Probability densities for the parameters in equation** (1)

| Parameter | Distribution |
|---|---|
| $R_*$ | log-uniform from from 1 to 100 |
| $f_p$ | log-uniform from 0.1 to 1 |
| $n_e$ | log-uniform from 0.1 to 1 |
| $f_l$ | log-normal rate, described in paper [9] |
| $f_i$ | log-uniform from 0.001 to 1 |
| $f_c$ | log-uniform from 0.01 to 1 |
| $N$ | point values: 1 to 10 000 |

From the equation we can compute $N$, which is the number of detectable civilizations, or longevity $L$:

$$L = \frac{N}{R_* f_p n_e f_l f_i f_c} \tag{2}$$

with parameters, i.e. probability densities and limits from Table 1. As Sandberg suggests, all distributions used in this model were log-uniform.

## 3.2 RARE EARTH MODEL

The Rare Earth model is based on the "rare Earth" theory that assumes that Earth is a very unique planet evolved under rare circumstances. This theory introduces equation:

$$N = N^* n_g f_p f_{pm} f_i f_c f_l f_m f_j f_{me} \tag{3}$$

We combined equation (3) with Drake's equation and used probability distributions from Tables 1 and 2. This instantly rules out the need of the $f_p$ (the fraction of stars with planets) parameter. Furthermore, product $f_l * f_i * f_c$ from Drake is equal to $f_i * f_c * f_l$ from Rare Earth, which gives us the final equation:

$$L = \frac{N^* n_g f_{pm} f_m f_j f_{me}}{R_* n_e} \tag{4}$$

and some new parameters:

- $N^*$ is the number of stars in the Milky Way galaxy (between 250 and 500 billion),
- $n_g$
- $f_{pm}$ is the fraction of planets that are metal-rich (between 1 and 10 percent),
- $f_m$ is the fraction of planets with a large moon (between 0.3 and 3 percent),
- $f_j$ is the fraction of solar systems with Jupiter-size planets (between 5 and 10 percent),
- $f_{me}$ is the fraction of planets with a critically low number of extinction events (between 1 and 10 percent).

In the Rare Earth model we also used log-uniform distribution, in order to compare it to the Sandberg model results.

# 4 EXPERIMENTS

## 4.1 Issues with log-uniform distribution

In analysing the two models, we focused primarily on how different distributions affect the results. Due to the shape of log-uniform distribution (see Figure 2), the part of the graph that is very close to zero has a significant impact on the final result. Since we have a logarithmic scale, the part from zero to one on the logarithmic scale corresponds to the range from zero to one percent, while the part from one to two percents corresponds to the range between one and one hundred percent, even though

**Table 2: Probability densities for the parameters in equation** (4)

| Parameter | Distribution |
|---|---|
| $N^*$ | log-uniform from 10.7 to 12.7 |
| $n_g$ | log-uniform from -1.3 to -0.8 |
| $f_{pm}$ | log-uniform from -3 to -0.7 |
| $f_m$ | log-uniform from -2.5 to -1.5 |
| $f_j$ | log-uniform from -1 to 0 |
| $f_{me}$ | log-uniform from -2.5 to -1.5 |

they appear to have the same weight on the logarithmic scale. The high values near zero therefore make it very sensitive to changes in parameter ranges and can even cause numerical errors when multiplications occur or at least strongly influence the final result.

For this reason, distributions whose values are close to zero at the boundaries of the parameter range are more stable with respect to changes in the parameters. We compared the stability of the log-uniform distribution with the log-normal distribution by slightly changing the lower bound of some parameters and observing the corresponding change in the distribution. The results in Figures 1 and 2, and later 3 and 4 indicate that the change of log-uniform distribution is much larger than that of log-normal distribution. Therefore, the log-normal distribution is much less dependent on the choice of the parameter range.

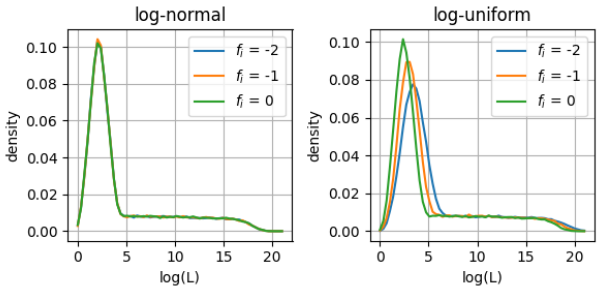



**Figure 1: Change of probability distribution with respect to change of lower range limit of parameter $f_i$.**

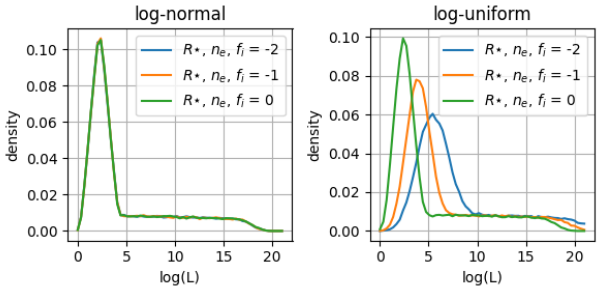



**Figure 2: Change of probability distribution with respect to change of lower range limit of parameters $R*$, $n_e$ and $f_i$.**

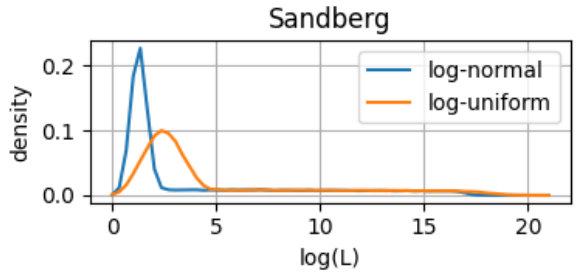



**Figure 3: Difference between log-uniform and log-normal distribution in the Sandberg model.**

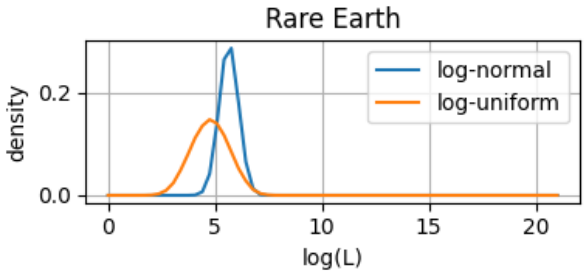



**Figure 4: Difference between log-uniform and log-normal distribution in the Rare Earth model.**

## 4.2 Parameter importance

In order to analyse the stability/sensitivity of the two models, we studied which parameters have the greatest impact on the final result. For this purpose, a dataset with different values and distributions for the parameters was created for the two models. Then, three subsets were taken, each containing only the subset with rows for which the probability that we survive at least L years is above 90%. The L options chosen were: 1000, 10 000, 100 000. The importance of the features in each of the subsets was then calculated using the Gini importance method implemented in the Python's scikit-learn decision tree regressor algorithm [7]. The feature importance scores are shown in Figures 5 to 10.

We found that in the Sandberg model, parameters 2 and 9 play the most important role, as you can see in Figures 6, 8 and 10, which show the importance of the parameters in calculating the probability that we survive 1000, 10 000 and 100 000 years.

In the model Rare Earth, on the other hand, parameters 5 and 7 are crucial for the prediction. This can be seen from Figures 5, 7 and 9, which show the importance scores of the parameters when calculating the same probabilities with the model Rare Earth.

## 5 DISCUSSION AND CONCLUSION

This research took two promising models from our earlier study [4] and analysed stability and sensitivity of the models and parameters. We analysed the stability of the log-uniform distribution compared to the log-normal distribution. To determine the difference between the two, Figures 1 and 2 are visually informative: changing the parameter range significantly affects the log-normal distribution, while the log-normal distribution is insensitive to these changes. Therefore, the log-normal distribution provides more reliable results, while the log-uniform distribution may cause some numerical curiosities. It seems reasonable to use distributions that rely mainly on the central values rather than the marginal values.

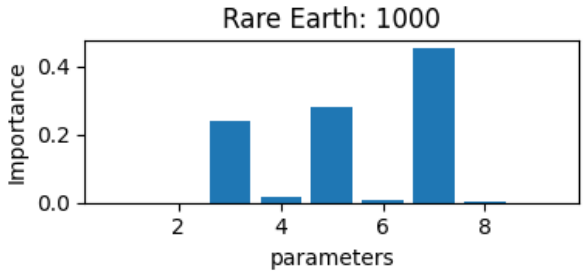



**Figure 5: Importance of parameters in Rare Earth model for estimating probability of surviving 1000 years.**

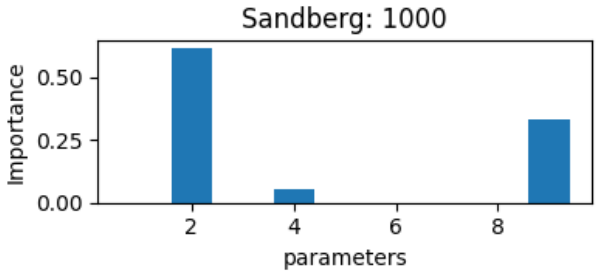



**Figure 6: Importance of parameters in Sandberg model for estimating probability of surviving 1000 years.**

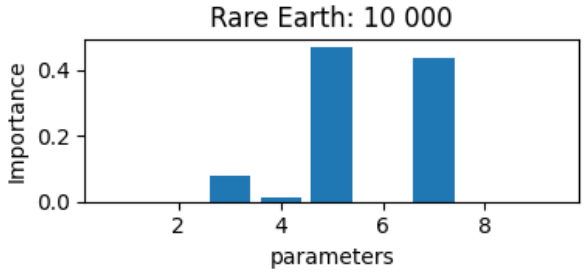



**Figure 7: Importance of parameters in Rare Earth model for estimating probability of surviving 10 000 years.**

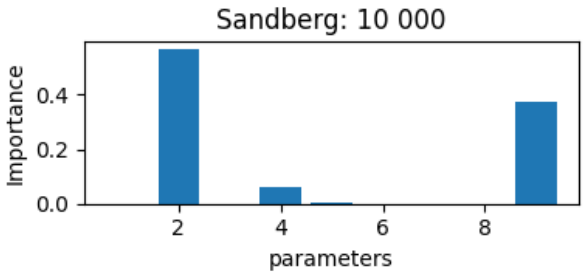



**Figure 8: Importance of parameters in Sandberg model for estimating probability of surviving 10 000 years.**

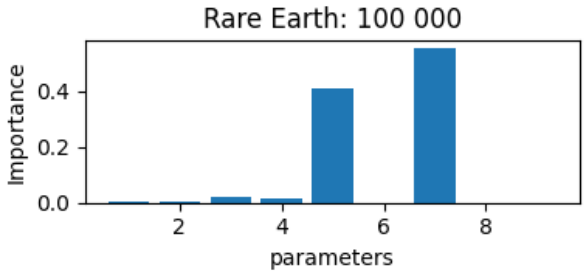



**Figure 9: Importance of parameters in Rare Earth model for estimating probability of surviving 100 000 years.**

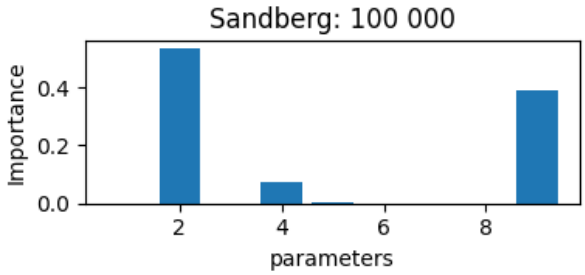



**Figure 10: Importance of parameters in Sandberg model for estimating probability of surviving 100 000 years.**

From Figures 3 and 4 we can observe that the Rare Earth model is considerably more optimistic than the Sandberg model. If we assume that Earth is very unique in our galaxy, we have the highest probability of living around 1 000 000 years. On the other hand, universe observations do not support well the uniqueness of our planet in terms of the large amount of suns with their planets. Further galaxy observations should provide more information which model fits the reality better.

From Figures 5 to 10, we can interpret that parameters 2, 5, 7, and 9 play the most important role in predicting the extinction of humanity. This seems novel compared to previous studies, and enables further discussion and studies regarding the causes and consequences of it. Whatever the case, while parameters seem to have numerically equal role and weight, studies of numerical relevance of the parameters of the equations (2) or (4) indicate significant differences.

Parameter 9 represents the choice of the distribution of the parameters. This is consistent with the distribution studies in this paper indicating that the probability curve for the longevity of human civilization strongly influences the obtained results.

Finally, while models do perform differently given different values of parameters, some patterns seem to emerge quite consistently if the parameters are set reasonably.

## ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0209).

## REFERENCES

[1] Frank Drake. 2015. *The Drake Equation: Estimating the Prevalence of Extraterrestrial Life Through the Ages*. Cambridge University Press.

[2] Marko Horvat. 2007. Calculating the probability of detecting radio signals from alien civilizations. *arXiv:0707.0011 [physics.pop-ph]*.

[3] Mansavi Lingam and Abraham Loeb. 2018. Relative likelihood of success in the searches for primitive versus intelligent extraterrestrial life. *arXiv:1807.08879 [physics.pop-ph]*.

[4] Anže Marinko, Klara Golob, Ema Jemec, Urša Klun, and Matjaž Gams. 2020. A new study of expected human longevity. *Informacijska družba, volume B*.

[5] Jason G. Matheny. 2007. Reducing the risk of human extinction. *Risk Analysis*, 27, 5, 1337.

[6] Jurij Nastran, Beno Šircelj, Drago Bokal, and Matjaž Gams. 2019. Sensitivity analysis of computational models that dissolve the fermi paradox. *Informacijska družba, volume A*.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[8] Anders Sandberg, Eric Drexler, and Toby Ord. 2018. Dissolving the Fermi paradox. *arXiv preprint arXiv:1806.02404*.

[9] Andrew E. Snyder-Beattie, Toby Ord, and Michael B. Bonsall. 2018. An upper bound for the background rate of human extinction. *Scientific Reports 9:11054. doi: 10.1038/s41598-019-47540-7*.

[10] Robert Strom G. 2015. We are not alone: extraterrestrial technological life in our galaxy. *Astrobiol Outreach 3: 144. doi: 10.4172/2332-2519.1000144*.

[11] Chenoa Tremblay and Steven Tingay. 2020. A seti survey of the vela region using the murchison widefield array: orders of magnitude expansion in search space. *doi: 10.1017/pasa.2020.27*.

[12] Peter Douglas Ward and Donald Eugene Brownlee. 2000. *Rare Earth: Why Complex Life Is Uncommon in the Universe*. Copernicus.

[13] Tom Westby and Christopher J Conselice. 2020. The astrobiological copernican weak and strong limits for intelligent life. *The Astrophysical Journal 896(1):58. doi: 10.3847/1538-4357/ab8225*.

[14] Beno Šircelj, Laura Guzelj Blatnik, Ajda Zavrtanik Drglin, and Matjaž Gams. 2019. Expected human longevity. *Informacijska družba, volume B*.

# Change ahead!
# Questioning and changing beliefs in online discussions

Lenart Motnikar†
Complexity Science Hub Vienna
lenart.motnikar@ait.ac.at

David Garcia
Complexity Science Hub Vienna
&
Graz University of Technology
dgarcia@tugraz.at

Hannah Metzler
Complexity Science Hub Vienna
&
Graz University of Technology
metzler@csh.ac.at

## ABSTRACT

Recent studies of persuasion and persuasibility in online discussions have predominantly focused on argument-specific features but not addressed extraneous factors that make someone question their beliefs in the first place. In this exploratory study, we sought to uncover factors underlying users' decisions to challenge their views in an online discussion forum and subsequently change them. We discovered that the examined psycholinguistic factors play a greater role in the questioning than the changing of opinions and further discuss the findings.

## KEYWORDS

persuasion, ChangeMyView, reddit, belief change

## 1 INTRODUCTION

Social media are becoming an increasingly dominant means of exerting persuasive influence on people. However, if not done appropriately and targeted at individuals who are not susceptible in the first place, attempts at persuasion can result in backfiring, pushing people further apart [1]. As these phenomena propagate through the population, affecting and changing society at large, persuasion in online social spaces has become an important topic of scientific inquiry.

Providing an open-access, natural discursive environment with user-labeled data, the *Change My View* (CMV) Reddit forum has become a popular research subject, being investigated in at least 20 studies [2] from fields like computational linguistics, behavioral design, and discourse studies.

On the forum, users write about their views on various topics with the purpose of having their views challenged. Users can then award the arguments of others with a "delta" if they succeed in changing their initial stance.

Studies of persuasion on the forum have mostly focused on what makes an argument persuasive and, to a lesser extent, what makes the users persuadable. The studies that inspected the latter mainly focused on features of the argument itself, measuring factors like linguistic, stylistic, and topical composition, as well as user interaction [3, 4, 5]. These studies, however, all focused on features pertaining directly to the arguments, neglecting a domain of potential explanatory significance – how users behave outside the argument.

Research in computational social science has indeed shown that the behavioral and linguistic traces of online activity can carry important information about the psychology of humans and the interactions between them [6].

Observing those would enable not only a deeper understanding of susceptibilities to being persuaded once a view has been questioned but also delving into the factors that influence the questioning of one's view in the first place.

Reddit provides a unique opportunity for such investigation, as each user's history of activity is publicly available and, because of the variety of discussion communities, less dependent on topic of discussion.

That being said, despite CMV's credo stating that the forum is

> "A place to post an opinion you accept may be flawed, in an effort to understand other perspectives on the issue."

only a small minority (13%) of the community's members ever post submissions on their own opinions, while the majority only participate in the discussions of others' views. While posting on CMV does not guarantee that a person is, in fact, open to view-change and the environment is not the only one where the process takes place, the relatively small share of submitting users implies that deliberately and openly challenging one's view is a relatively unique phenomenon, even within a purposed community like CMV.

To fill the identified gap in current research on persuasion, we set out to explore the factors associated with users' decisions to, first, challenge their opinions on CMV, and second, to end up changing them.

To answer these questions, we inspected the users' activities on Reddit before they joined the CMV community. As this is an exploratory endeavor without much theoretical foundation, we focused on surface-level parameters, observing the user's posting patterns, stylistic and linguistic features, indicators of personality, and their community affiliation.


*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*

## 2 METHOD

We collected submissions and comments that were posted to CMV between January 1st and December 31st, 2020, excluding those that were removed by moderators, or made by bots and deleted accounts. This left us with 31,419 submissions and 1,563,865 comments, authored by 158,724 unique users, 21,168 of whom posted submissions.

We studied users who made their first contribution to CMV in the studied period and were active on the forum over a span of at least seven days. While this threshold is somewhat arbitrary, it allowed us to exclude users who were mere passersby of the community (who may be unfamiliar, unserious, or even purposefully disruptive), while retaining a representative sample accounting for a majority (69%) of newcomer-created content.

We then downloaded the users' post histories one year before their first post (submission or comment) on CMV and imposed additional filters, keeping the users who:

a) made less than 10,000 submissions and comments, to exclude potential bots and spammers, and
b) made at least 10 posts containing 100 analyzable tokens before joining CMV, to ensure enough data.

For each user, we created two separate corpora, one of pre-CMV submissions and one of pre-CMV comments. We then analyzed their posts across various domains, excluding deleted and non-English (estimated automatically, using [7]) posts from text analysis.

### 2.1 Investigated features

**Posting behavior.** First, we collected data on the users' posting behavior, including days of activity pre-CMV activity, the number of communities they were involved with, the average length of submissions and comments, and their daily rates of posting.

**Psycholinguistic characteristics.** Second, we scored the post histories on selected categories of the LIWC2015 dictionary [8], a popular tool for psycholinguistic research, containing common words and word stems categorized by grammatical and semantic categories.

We selected features relating to grammar, as well as selected psychological categories. The latter included affective, cognitive, social, perceptual, and biological processes, drives, relativity, and time orientations.

**Formatting and structure.** Third, we looked at the outward appearance and structure of users' posts by extracting Markdown formatting features, namely the use of bold, italics, quotations, links, and un/ordered lists.

**Personality.** Fourth, we built a predictor of BIG5 personality traits by matching the top and bottom 100 n-grams that were shown to be associated with each personality dimension in [9] and summing their correlation-weighted scores.

**Reddit communities.** In addition, we also explored differences in the communities where the users were active, to see if particular communities are more or less popular within a certain population. We looked at the subreddits where the users posted and calculated the percentages of affiliated users in the studied groups.

We then set to explore the data in two problems, comparing two sets of users in each task.

### 2.2 Task 1: Questioning one's view

In Task 1, we explored the characteristics of CMV users who posted submissions questioning their views by comparing them to those who only commented on others' posts but never submitted posts on their own views.

After filtering by the previously mentioned criteria, the experimental group consisted of 4,639 users who posted at least one submission on CMV.

We compared those users to a control group of the same size, randomly selected from the users that passed the criteria but never posted their submission (although they may have done so after the studied period). From here on, we refer to these groups as questioning (Q) and non-questioning (Non-Q).

### 2.3 Task 2: Changing one's view

In the second part, we were interested in finding the characteristics underpinning one's susceptibility to view-change. For this, we divided questioning users into two subgroups: those susceptible (S) and non-susceptible (Non-S) to view-change.

We deemed a submission as ending in view-change if its author has awarded a "delta" that has been confirmed by the forum's Delta-Bot, which checks for rule compliance.

We selected CMV submissions that garnered at least 10 comments (indicating that some discussion took place) and compare authors who changed their views in either 100% (n=1,435) or 0% (n=1,204) of the submissions they posted. We opted for this distinction following [3], presupposing that the differences would be more notable between extremes.

## 3 RESULTS

In both tasks, we conducted a series of Bonferroni-Holm-corrected significance tests, comparing the features of the users' pre-CMV submission and comment corpora separately. We present results in Table 1 for Task 1 and Table 2 for Task 2, showing only features that yielded significant differences, due to spatial limitations.

In Task 1, we observed that questioning users, on average, posted submissions more often while having a shorter duration of pre-CMV activity.

Regarding LIWC, users differed in most of the studied categories. In most cases, the trend pointed in the same direction in both submissions and comments. In some cases, the difference was significant only in one set, and in a few, the trends in submissions and comments opposed one another.

Regarding formatting, questioning users used more ordered lists in both sets of corpora, while they used fewer quotes in the comments.

The users' posts exhibited quite inconsistent manifestations of personality, expressing lower neuroticism in submissions, while in the comments, they showed higher agreeableness, extraversion and conscientiousness, and lower openness.

**Table 1. Significance testing results in Task 1.** The numbers represent effect sizes (Cohen's *d*). Arrow direction represent how the feature expresses in Q users relative to Non-Q. The number of arrows denotes significance at p<.05, p<.01, p<.001, or p<.0001.

| Feature | Characteristic of questioning? | | | |
|---|---|---|---|---|
| **Posting features** | | | | |
| Submissions per day | .32 | ↑↑↑↑ | | |
| Days of activity | -.39 | ↓↓↓↓ | | |
| | *Submissions* | | *Comments* | |
| **LIWC** | | | | |
| Function words | .14 | ↑↑↑↑ | .07 | ↑ |
| Pronouns | .16 | ↑↑↑↑ | .21 | ↑↑↑↑ |
| Personal pronouns | .10 | ↑↑↑ | .22 | ↑↑↑↑ |
| 1st person singular | -.03 | | .31 | ↑↑↑↑ |
| 1st person plural | -.10 | ↓↓ | .00 | |
| 2nd person | .14 | ↑↑↑↑ | .04 | |
| 3rd person plural | .02 | | -.11 | ↓↓↓↓ |
| Impersonal | .15 | ↑↑↑↑ | .06 | ↑ |
| Articles | -.06 | | -.21 | ↓↓↓↓ |
| Prepositions | -.07 | | -.16 | ↓↓↓↓ |
| Common adverbs | .11 | ↑↑↑↑ | .08 | ↑↑ |
| Conjunctions | .13 | ↑↑↑↑ | .11 | ↑↑↑↑ |
| Common adjectives | .10 | ↑↑↑ | -.08 | ↓↓ |
| Comparisons | .12 | ↑↑↑↑ | -.04 | |
| Interrogatives | .26 | ↑↑↑↑ | .12 | ↑↑↑↑ |
| Numbers | -.15 | ↓↓↓↓ | -.08 | ↓↓ |
| Quantifiers | -.03 | | -.08 | ↓↓ |
| Positive emotion | -.01 | | .11 | ↑↑↑↑ |
| Negative emotion | .16 | ↑↑↑↑ | -.03 | |
| Social processes | .24 | ↑↑↑↑ | .04 | |
| Cognitive processes | .16 | ↑↑↑↑ | .12 | ↑↑↑↑ |
| Perceptual processes | -.04 | | .09 | ↑↑↑ |
| Drives | .03 | | -.08 | ↓↓ |
| Present focus | .08 | ↑ | .08 | ↑↑ |
| Relativity | -.20 | ↓↓↓↓ | -.17 | ↓↓↓↓ |
| **Formatting** | | | | |
| Quote | .02 | | -.08 | ↓↓ |
| Ordered list | .07 | ↑ | .08 | ↑↑ |
| **Personality** | | | | |
| Openness | -.06 | | -.16 | ↓↓↓↓ |
| Conscientiousness | -.04 | | .08 | ↑↑ |
| Extraversion | .04 | | .16 | ↑↑↑↑ |
| Agreeableness | -.06 | | .09 | ↑↑↑ |
| Neuroticism | -.15 | ↓↓↓↓ | .03 | |

In Task 2, there were fewer differences compared to Task 1. Regarding posting features, susceptible users exhibited a lower rate of posting submissions. There were no observable differences in LIWC categories, while in formatting, susceptible users exhibited a slightly higher use of ordered lists in the comments. Regarding personality, susceptible users expressed higher agreeableness and neuroticism in the comments.

We also inspected if the user groups in both tasks differ in the communities they contribute to. Table 3 presents ratios between the percentages of users who were affiliated with the community in each group, with a bottom threshold of 2%.

**Table 2. Significance testing results in Task 2.** Arrow directions represent feature expression in S users relative to Non-S.

| Feature | Characteristic of susceptibility? | | | |
|---|---|---|---|---|
| **Posting features** | | | | |
| Submissions per day | -.17 | ↓↓ | | |
| | *Submissions* | | *Comments* | |
| **Formatting** | | | | |
| Ordered list | .08 | | .14 | ↑ |
| **Personality** | | | | |
| Agreeableness | .01 | | .15 | ↑ |
| Neuroticism | .11 | | .20 | ↑↑↑↑ |

In Task 1, for example, questioning users had a 2.37 times higher likelihood to post on *r/askphilosophy* (a forum for discussion of philosophical ideas) and a relative likelihood of 0.3 to post on *r/bestof* (a forum where users share their favorite comments across all Reddit). Similarly, in Task 2, susceptible users were 2.7 times more likely to post on *r/getdisciplined* (a support community for self-improvement) but had a likelihood of 0.58 to post on *r/socialism*.

**Table 3: Quotients of subreddit association rates between Q and Non-Q users in Task 1 and S and Non-S users in Task 2.**

| Task 1 | | Task 2 | |
|---|---|---|---|
| subreddit | ratio | subreddit | ratio |
| askphilosophy | 2.37 | getdisciplined | 2.70 |
| SuicideWatch | 2.08 | woooosh | 2.33 |
| FreeKarma4U | 1.96 | confidentlyincorrect | 2.31 |
| ask | 1.95 | ShitAmericansSay | 2.31 |
| findareddit | 1.90 | antimeme | 2.28 |
| … | | … | |
| The_Mueller | 0.36 | AbruptChaos | 0.62 |
| LeopardsAteMyFace | 0.35 | sports | 0.62 |
| MaliciousCompliance | 0.33 | PoliticalDiscussion | 0.61 |
| LivestreamFail | 0.33 | PS4 | 0.59 |
| bestof | 0.30 | socialism | 0.58 |

## 4 DISCUSSION

In this study, we sought to uncover parameters that might carry explanatory information about a user's tendency to openly question and then change their views. First, we compared users who posted submissions on CMV to those that only commented. Second, we compared the submitters who always ended up changing their views to those that never did.

We first observed that the users who posted submissions to CMV also had a higher rate of posting submissions elsewhere, before they joined the forum, indicating that the users who submit to CMV are in general more inclined to post submissions, which could be due to many factors. We observe a similar albeit weaker discrepancy in Task 2, where

a higher rate of posting submissions was characteristic of non-susceptible users.

We then noticed that the time of the questioning users' activity or Reddit before their first contribution to CMV was shorter on average. One explanation could be that the submissions were posted from secondary accounts, perhaps to anonymize one's expression of a view they would not feel comfortable sharing otherwise. Despite our intentions to limit such "throwaway" accounts by imposing a limit of minimum activity, enough might have remained to have affected the data.

We further observed that questioning users have a significantly different linguistic profile, as significant differences appeared in several measured LIWC categories. Of those, function words and pronouns in particular have been studied the most and are known to bear psychological relevance, as they reveal the focus of the author's attention and the relations between the entities discussed [10]. Higher (personal) pronoun use, which was characteristic of questioning users, generally points towards more personal and people-oriented language. However, when it comes to interpretation, it is important to also consider the different contexts of submissions and comments, which differ in who they're directed to. In submissions, where users address a general audience, we observed that questioning users used more second person ("you") and less first-person plural ("we") pronouns. The role of second person has been predominantly studied in close relationships, where it is likely to entail confrontation [10]. However, in the context of submissions, this is not likely to be the case. As they are directed towards an unspecified reader, it is probably more likely that the use of "you" is meant in a manner that is inquisitive or directing (e.g., "What do you guys think?", "You should try this!"), showing initiative and an interest in others. This interpretation is also in line with the observation that questioning users used more interrogatives.

Next, the lesser use of first-person plural ("we") in submissions could indicate a lower degree of community affiliation and belonging. It has previously been suggested that binding one's view to a group disperses the feeling of responsibility for it [5]. If questioning users hold beliefs as their own rather than representing a group they identify with, they may be more likely to question their views.

In the comments, we observed two further pronoun-related trends. In particular, questioning users used more first-person singular ("I"), which entails greater self-focus, perhaps as a means of explaining oneself, and less third-person plural ("they"), indicating a lesser focus on an outgroup or people in general.

Furthermore, we observed differences in several other grammatic and semantic categories in both submissions and comments. These point towards thematic and topical discrepancy between the users' use of language. As a general observation, questioning users used fewer numbers, articles, prepositions, and relativity, which indicates a lesser propensity for complex, analytic, and concrete language. This is contrasted by a higher use of words in the psychological process categories, supporting the previous explanation that questioning users tend to be more personal in their expression.

This considered, it is important to note that the effect sizes of observed differences are minimal, and without a deeper examination of context, nuanced interpretation is difficult.

An interesting observation is that across all features, the groups differed a lot more in Task 1 than in Task 2. This shows that the psycholinguistic characteristics underpinning one's tendency to challenge their view on CMV play greater importance compared to the ones behind their susceptibility to award "deltas". At the same time, they show that the users who decide to submit to CMV might gravitate towards a certain type of user, begging the question of generalizability of studies of the forum.

The personality measures showed several differences in both tasks but were inconsistent when comparing expressions in submissions and comments. Given that differences for each dimension were shown only in one set of corpora, this might high indicate contextual dependency. Research has indeed shown that word correlation-based measures of personality depend on communication contexts [11], which could also apply to those of submitting and commenting. The second contextual consideration is that the tokens used for personality estimation were taken from a study of posts on Facebook and might therefore not translate well to the social environment on Reddit.

We also observed that certain subreddits were more or less likely to be visited by the studied groups, indicating some kind of community preferences, although it is not obvious what underlies them. Going forward, it would be interesting to examine if these differences are driven by topic or by specific social characteristics.

The main takeaway from this study is that the explored factors, particularly those regarding language, have a greater role in underlying questioning one's views on CMV, than changing them. However, as noted in the beginning, questioning users posted more submissions overall. It is important to note that although we interpreted our findings through the lens of questioning beliefs, this might not be the main explaining factor behind the observations. It could be that the differences we observed are driven more by this general propensity to post submissions than a wish to challenge one's views.

In the future, it would therefore be necessary to explore this question further. For example, one could investigate if similar differences exist between submitters and non-submitters in other communities or if these effects scale with the users' rates of posting submissions. To better understand the mechanisms behind challenging beliefs, we would have to control for such factors, as well as discern how motivations for submitting in general interact with those specifically relating to questioning views.

## ACKNOWLEGMENT

D. G. and H. M. acknowledge funding from the Vienna Science and Technology Fund through the project "Emotional Well-Being in the Digital Society" (Grant No. VRG16-005).

## REFERENCES

[1] Simon Schweighofer, Frank Schweitzer, and David Garcia. 2020. A Weighted Balance Model of Opinion Hyperpolarization. *Journal of Artificial Societies and Social Simulation* 23, 3 (2020), 5. DOI: https://doi.org/10.18564/jasss.4306.

[2] Nicolas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiersler and Michael Zimmer. 2021. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society* 7, 2 (Jun, 2021), 1-14. DOI: https://doi.org/10.1177/20563051211019004.

[3] Humphrey Mensah, Lu Xiao, and Sucheta Soundarajan. 2019. Characterizing Susceptible Users on Reddit's ChangeMyView. In *Proceedings of the 10th International Conference on Social Media and Society (SMSociety '19)*. Toronto, ON, Canada, 102-107. DOI: https://doi.org/10.1145/3328529.3328550.

[4] John Hunter Priniski and Zachary Horne. 2018. Attitude Change on Reddit's Change My View. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*. Madison, WI. 2279- 2284.

[5] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lilian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of the 25th international conference on world wide web*. Montreal, QC, Canada, 613-624. DOI: https://doi.org/10.1145/2872427.2883081.

[6] Michele Settanni, Danny Azucar, and Davide Marengo. 2018. Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking* 21, 4 (April, 2018), 217-228. DOI: https://doi.org/10.1089/cyber.2017.0384.

[7] Michal Mimino Danilak. 2021. langdetect 1.0.9. *Python Package Index (PyPI)*, https://pypi.org/project/langdetect/1.0.9/.

[8] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. University of Texas at Austin, Austin, TX.

[9] H. Andrew Schwartz et al. 2013. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PloS one 8*, 9 (2013), e73791. DOI: https://doi.org/10.1371/journal.pone.0073791.

[10] Yla E. Tausczik, and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology,* 29, 1 (March, 2010), 24-54.

[11] Matthias R. Mehl, Megan L. Robbins, and Shannon E. Holleran. 2012. How Taking a Word for a Word Can Be Problematic: Context-Dependent Linguistic Markers of Extraversion and Neuroticism. *Journal of Methods and Measurement in the Social Sciences* 3, 22 (2012). 30-50. DOI: https://doi.org/10.2458/v3i2.16477.

# Kaj se lahko naučimo od Jacques Mehlerja, klasičnega kognitivnega znanstvenika

## What can we Learn from Jacques Mehler, a Classical Cognitive Scientist

Amanda Saksida
Institute for Maternal and child health Burlo Garofolo – Trieste
Italija
amanda.saksida@icloud.com

**POVZETEK**

Prispevek prikazuje življenjsko delo Jacquesa Mehlerja, ki je bil eden uspešnejših evropskih raziskovalcev razvoja človeške kognicije, še posebej zgodnjega razvoja govora. Ob tem predstavi glavne predpostavke klasične kognitivne znanosti – modularnost uma ter vlogo narave in vzgoje pri razvoju in delovanju miselnih procesov – in opiše, katere vpoglede je omogočilo empirično raziskovanje teh predpostavk v preteklih desetletjih. Na kratko tudi oriše nova spoznanja, ki so kognitivno znanost v zadnjih dveh desetletjih dodobra spremenila in ki so deloma vplivala tudi na njegovo delo. Način, kako je Mehler ta nova spoznanja vedno znova integriral v svoje delo, lahko predstavlja enega od modelov sinteze empiričnega in teoretskega raziskovanja.

**KLJUČNE BESEDE**

klasična kognitivna znanost, modularnost uma, razvoj govora, Jacques Mehler

**ABSTRACT**

The article shows the life work of Jacques Mehler, who was one of the most successful European researchers in the field of the development of the human mind, especially early language acquisition. The article presents the main assumptions of classical cognitive science – the modularity of the mind and the role of nature and nurture in the development and functioning of the mind – and describes which insights have been enabled by Mehler's empirical research of these assumptions over the past decades. New findings are also briefly presented that have changed cognitive science over the last two decades and that have partly influenced his work. The way in which Mehler has repeatedly integrated these new insights into his work can represent one of the models of the synthesis of empirical and theoretical research.

**KEYWORDS**

classical cognitive science, modularity of mind, language acquisition, Jacques Mehler



### 1 Klasična kognitivna znanost in Mehlerjev doprinos

V letu 2020 je v Parizu v starosti 83 let po dolgi nevrodegenerativni bolezni umrl Jacques Mehler, eden izmed pomembnih mladih akterjev tako imenovane kognitivne revolucije, ki se je zgodila v 60-ih letih prejšnjega stoletja in je pomenila odmik od takrat prevladujočega behaviorizma k proučevanju vrojenih lastnosti kognicije. Od leta 1975 do 2001 je vodil psiholingvistični laboratorij v Parizu (Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-ENS). Zaradi po njegovem mnenju prezgodnje upokojitve v francoskem CNRS se je leta 2001 lotil še vzpostavitve laboratorija Language, Cognition and Development Lab na SISSA-ISAS v Trstu, ki ga je vodil do končne upokojitve leta 2016.

Mehler je kot direktor pariškega laboratorija veljal za klasičnega kognitivnega znanstvenika, ki je človeško kognicijo raziskoval v skladu z osnovnima predpostavkama, da je um modularen ter da je večina miselnih procesov vrojenih. Ideja o modularnosti uma se je deloma napajala iz raziskav zgodnje nevrologije, vendar pa jo je v drugi polovici 20. stoletja najbolje izpeljal Jerry Fodor. Fodorjeva različica teorije o modularnosti uma, ki jo je povzel tudi Mehler, ne nudi neposredne navezave na fiziološke procese, predpostavlja pa, da na vsakem področju (modulu) uma/kognicije veljajo drugačni načini učenja in zaznavanja (angl. domain specificity), ki niso neposredno vezani na drug modul (angl. information encapsulation) in ki niso nujno vezani na eno samo čutilo [1]. Ideja o vrojenosti miselnih procesov je, podobno, izhajala iz spoznanja o visoki specializaciji nekaterih delov kognicije že zelo zgodaj v razvoju, še najbolj izrazito v razvoju govora [2]. Skladno s to idejo je učenje pravzaprav zgolj sprožanje nastavitev parametrov, ki so sami po sebi vrojeni [3] Kognitivna znanost, ki je predpostavljala modularnost uma in vrojenost miselnih procesov je pomenila neposredno kritiko behaviorizma, ki je predpostavljal splošne mehanizme učenja za vse miselne procese in po katerem je učenje vedno neposredni odziv na zunanje dražljaje [4]. Razprava o vlogi narave in vzgoje je sicer stara tisočletja, in mnenja o tem, da so nekateri miselni procesi vrojeni, saj jih lahko opazujemo takoj po rojstvu ali še

pred njim, se še danes silovito krešejo z mnenji, da so ti procesi posledica učinkovitih splošnih učnih mehanizmov.

V teh teoretskih okvirih je Mehler izpeljal vrsto empiričnih raziskav o tem, kako je človeško zaznavanje selektivno in pogojeno z vrojenim znanjem tudi na področju prepoznavanja in učenja maternega jezika. Ugotovil je, da je zlog osnovna zaznavna enota v govoru in da je prepoznava zloga kot osnovne zaznavne enote pomembna pri učenju in segmentaciji besed [5], [6], in to že od rojstva naprej [7]. Vendar pa so že novorojenčki pozorni tudi na druge pomembne elemente govora, kot so premori in spremembe v intonaciji [8], [9]. Skupaj s študenti je raziskoval zmožnost razločevanja različnih jezikov ob rojstvu in ugotovil, da novorojenčki prepoznajo materin glas ter ritem jezika, ki so ga poslušali že pred rojstvom, ter ga ločijo od jezika z drugačnim ritmom, vendar pa ne ločijo dveh ritmično podobnih jezikov [10], [11]. Kljub določeni meri skepse glede neposredne povezave med (vrojenimi) miselnimi procesi in njihovo fiziološko (nevrološko) podlago je bil zavezan eksperimentalnemu delu ter sodelovanju pri metodoloških inovacijah, potrebnih za raziskave zgodnjega razvoja. To je kasneje omogočilo tudi nekatera dognanja s področja nevrologije kognitivnih procesov, ki jih je preučeval. Med drugim je prvi uporabil NIRS (angl. near-infrared spectroscopy) tehniko optične topografije pri novorojenčkih ter tako prvi pokazal, da človek že ob rojstvu procesira govor v levi možganski polovici [12].

## 2 Mehlerjeva integracija novih idej v klasično kognitivno znanost

Kognitivna znanost se je na prelomu tisočletja zopet začela korenito spreminjati. Bolj množično so se začele zbujati kritike teorije o modularnosti uma ter selektivnih zaznavnih in učnih mehanizmov. Naraslo je tudi zanimanje za vlogo splošnih statističnih učnih mehanizmov pri učenju govora, na primer zaznavanja pogostosti pojavitve osnovnih gradnikov jezika, fonemov, zlogov, besed, ter pogojnih verjetnosti sopojavljanja teh gradnikov v jeziku [13], [14]. To je po naključju sovpadlo tudi z Mehlerjevim premikom iz Pariza v Trst leta 2001. Novi laboratorij v Trstu se je začel ukvarjati z odnosom med statističnim učenjem in osnovnimi predpostavkami klasične kognitivne znanosti. S skupino mladih sodelavcev je Mehler preučeval lastnosti in omejitve statističnega učenja pri segmentaciji in učenju besed. Statistično učenje recimo deluje drugače na samoglasnikih kot na soglasnikih [15], [16], kadar pa so si statistične in prozodične informacije v nasprotju, se človeški um bolj zanaša na prozodične [17]–[19].

Opažanje, da je zaznavanje selektivno, je pripeljalo tudi do študij bolj ali manj specializiranih mehanizmov zaznavanja, npr. zaznavanje identitete (ponavljanja, npr. ponavljanja zlogov) in zaznavanje robov (npr. boljše pomnjenje zlogov na robovih besed), ki v veliki meri olajšajo zgodnje učenje jezika [20]–[22]. Obenem pa so v laboratoriju potekale tudi raziskave o tem, kako razvoj govora, kot specializiranega znanja, vpliva na druge dele človeške kognicije, na primer na centralne nadzorne in izvršilne funkcije. Na primer, vsakodnevno poslušanje dveh ali več jezikov vpliva na izvršilne funkcije že kmalu po rojstvu: dojenčki iz dvojezičnih družin že pri 7 mesecih izkazujejo boljšo kontrolo in inhibicijo kot njihovi enojezični vrstniki [23], [24].

Mehlerjeva izhodiščna pozicija je bila torej jasna in večina objavljenih del se je ukvarjala z omejitvami splošnih učnih mehanizmov ter visoko specializiranimi mehanizmi, ki so po njegovem prepričanju najverjetneje vrojeni (specializirani mehanizmi zaznavanja, stavčni ritem in prozodija, soglasniki-samoglasniki). Vendar pa je pri svojem delu ostajal trdno zavezan empiričnemu preverjanju glavnih teoretskih vprašanj s pomočjo čim bolj objektivnega in nepristranskega opazovanja človeških odzivov od rojstva naprej, pravzaprav podobno kot Piaget, čeprav so ju ločevala nesoglasja. Ker mu je empirično raziskovanje omogočalo vsaj delno distanco od teoretskega dela, ostaja odprto vprašanje, kako bi na razvoj kognitivne znanosti gledal danes.

## 3 Kognitivna znanost danes v odnosu do Mehlerjevega dela

Predstavljena teoretska vprašanja kognitivne znanosti so bila v zadnjih letih soočena z novimi podatki, ki so kazali na to, da lahko splošni kognitivni primanjkljaji zaradi spremenjenega vnosa podatkov pripeljejo do specifičnih razvojnih motenj. Na primer, specifična jezikovna motnja bi bila lahko posledica centralnega primanjkljaja v procesiranju hitrih zvočnih dražljajev [25]. Podobno sosledje morda velja tudi za disleksijo [26], [27]. Vendar pa mnenja o izvoru učnih razvojnih motenj ostajajo deljena in zato še vedno prevladujejo kognitivni modeli, ki predvidevajo modularnost posameznih področij kognicije [28], [29].

Ker so kognitivni procesi nujno posledica dejavnosti možganov, ideja modularnosti uma tudi v svojih novejših različicah vselej predpostavlja, da so specializirani procesi tisti, ki zasedajo nek točno določen predel možganskega tkiva [30]. To idejo so nedavna spoznanja v nevroznanosti dodobra načela z dokazi, da so posamezni možganski moduli, ki so bili tradicionalno razumljeni kot osnovni kognitivni moduli, v resnici deli nevronskih mrež, ki pa so v možganih pogosto uporabljene večkrat in za različne namene (angl. neural reuse, neural redeployment) [31], [32]. Še več, bistvo specializacije nevronskih mrež verjetno ni v njenih osnovnih gradnikih, možganskih modulih, temveč v načinu, kako so ti gradniki povezani. Zato je mogoče za iste kognitivne funkcije opazovati dejavnost različnih nevronskih mrež, ali pa obratno, dejavnost istih (ali vsaj navidezno istih) nevronskih mrež za različne kognitivne funkcije [33]. Primer za slednje so ekspertne veščine, ki jih eksperti lahko navidezno opravljajo avtomatizirano, vendar pa obenem ohranjajo centralni nadzor nad dinamiko dogajanja, kar bi lahko nakazovalo, da je za dva procesa odgovorno eno (ali vsaj na videz eno) nevronsko omrežje [34].

Čeprav so se kognitivni modeli delovanja kognicije v preteklosti lahko ogradili od modelov nevrološkega delovanja, ker ti niso bili v neposrednem nasprotju s prvimi, ima ponujeni model organizacije nevronskih mrež neposredne posledice tudi za kognitivne modele, saj predpostavlja, da so vsa specializirana znanja modularna samo v zelo abstraktnem

smislu, ter da so nujno posledica učenja in ne vrojena. Vendar pa obenem ponudi svežo rešitev uganke, s katero se že dolgo soočajo raziskovalci specifičnih razvojnih motenj, ki se jim izmika enoznačna razlaga izvora teh motenj. Možno je namreč, da kognitivni profili in razvojne motnje niso posledica lastnosti in pomanjkljivosti v posameznih možganskih modulih, temveč predvsem načina, kako so organizirane nevronske mreže [35], [36]. Organizacija nevronskih mrež pa je v veliki meri odvisna od dogodkov v času nastajanja človeškega bitja.

To pa je pravzaprav pot, ki ji je sledil tudi Mehler, ko je zametke razvoja govora iskal in razpoznaval v obdobju globoko pred prvo besedo, že takoj po rojstvu. Na novorojena človeška bitja je vedno gledal kot na aktivne, zavedajoče se soudeležence pri lastnem razvoju, in logična posledica tega pogleda je bila, da so se nekateri njegovi študentje in sodelavci lahko spustili na področje raziskovanja izkušenj in znanj, ki jih zarodki pridobijo že pred rojstvom. Nove raziskave tako med drugim ugotavljajo, kako lahko pri zarodkih merimo in spodbujamo njihovo zmožnost slušne (glasba, govor, glas) ali vidne prepoznave (obrazne poteze) ter pomnjenja in kako lahko to učinkuje na organizacijo nevronskih mrež že pred rojstvom [37], [38]. In tako se nadaljuje naloga, ki si jo je zadal Mehler: ugotoviti, koliko lahko prispeva dejavnost in stimulacija na zmožnost zaznavanja in razločevanja ter na učenje, vendar pa ne več pri novorojenčkih, kot je to počel on, temveč že pred rojstvom.

## 4 Sklep

Jacques Mehler je svoje področje zapustil v času, ko je gotovosti v zvezi z razumevanjem kognicije na videz manj, saj so se zrahljali klasični kognitivni modeli. Vendar pa se zdi, da so nedavna spoznanja o povezljivosti možganov odprla nove možnosti za razumevanje razvoja in delovanja uma. In prav mogoče je, da bi se tudi Jacquesovo delo, če bi bil še vedno dejaven, usmerilo v raziskovanje nevronskih omrežij, ki sodelujejo pri procesiranju jezika od rojstva naprej ali pa še pred rojstvom. Gotovo pa je, da bi ga radovednost in natančnost, ki ju je gojil pri svojem delu, še naprej vodila v tehtno pretresanje mej ter omejitev modelov razvoja in delovanja človeškega uma.

In prav to je vodilo, ki je lahko koristno za vsakogar, ki ga zanima razvoj človeškega uma. Z natančnim pretresanjem možnosti, ki jih odpira vsak model delovanja človeškega uma, in možnih odgovorov, ki jih nudijo človeški odzivi na dražljaje, lahko vsakdo od nas prispeva delež novega vedenja o pomenu in funkciji modulov – vrojenih ali priučenih, anatomskih ali kognitivnih – ki omogočajo specializirana znanja, lastna človeku.

## ZAHVALA

Za pomoč pri zbiranju informacij za prispevek se zahvaljujem kolegom, ki so v istem času pisali retrospektivne članke o življenju in delu Jacquesa Mehlerja: Jean Remy Hochmann, Judit Gervain, Agnes Kovacs, Stanislas Dehaene.

## REFERENCES

[1] J. A. Fodor, *The modularity of mind*. MIT Press, 1983.
[2] N. Chomsky, „A Review of B . F . Skinner ' s Verbal Behavior“, *Language (Baltim).*, let. 1, str. 26–58, 1959.
[3] E. Gibson in K. Wexler, „Triggers“, *Linguist. Inq.*, let. 25, št. 3, str. 407–454, 1994.
[4] B. F. Skinner, „Behaviorism at Fifty“, *Science*, let. 140, št. 3570, str. 951–958, 1963.
[5] A. Cutler in J. Mehler, „The periodicity bias“, *J. Phon.*, let. 21, št. 1/2, str. 103–108, 1993.
[6] J. Mehler, J. Y. Dommergues, U. Frauenfelder, in J. Segui, „The Syllable's Role in Speech Segmentation“, *J. verbal Learn. evrbal Behav.*, let. 20, str. 298–305, 1981.
[7] J. Bertoncini in J. Mehler, „Syllables as Units in InfantSpeech Perception“, *Infant Behav. Dev.*, let. 4, str. 247–260, 1981.
[8] J. Bertoncini, C. Floccia, T. Nazzi, in J. Mehler, „Morae and syllables: rhythmical basis of speech representations in neonates.“, *Lang. Speech*, let. 38, št. 4, str. 311–329, 1995.
[9] A. Christophe, J. Mehler, in N. Sebastián-Gallés, „Perception of Prosodic Boundary Correlates by Newborn Infants“, *Infancy*, let. 2, št. 3, str. 385–394, jul. 2001.
[10] F. Ramus, M. D. Hauser, C. Miller, D. Morris, in J. Mehler, „Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys“, *Science*, let. 288, št. 5464, str. 349–351, apr. 2000.
[11] T. Nazzi, J. Bertoncini, in J. Mehler, „Language discrimination by newborns: toward an understanding of the role of rhythm.“, *J. Exp. Psychol. Hum. Percept. Perform.*, let. 24, št. 3, str. 756–66, jun. 1998.
[12] M. Peña *idr.*, „Sounds and silence: An optical topography study of language recognition at birth“, *Proc. Natl. Acad. Sci.*, let. 100, št. 20, str. 11702–11705, 2003.
[13] J. R. Saffran, R. N. Aslin, in E. L. Newport, „Statistical Learning by 8-Month-Old Infants“, *Science*, let. 274, št. 5294, str. 1926–1928, 1996.
[14] D. Swingley, „Statistical clustering and the contents of the infant vocabulary“, *Cogn. Psychol.*, let. 50, št. 1, str. 86–132, feb. 2005.
[15] L. L. Bonatti, M. Peña, M. Nespor, in J. Mehler, „On consonants, vowels, chickens, and eggs.“, *Psychol. Sci.*, let. 18, št. 10, str. 924–5, okt. 2007.
[16] J. M. Toro, M. Nespor, J. Mehler, in L. L. Bonatti, „Finding words and rules in a speech stream: functional differences between vowels and consonants.“, *Psychol. Sci.*, let. 19, št. 2, str. 137–44, feb. 2008.
[17] A. L. Ferry *idr.*, „On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain“, *Dev. Sci.*, let. 19, št. 3, str. 488–503, 2016.
[18] A. Saksida, A. Langus, in M. Nespor, „Co-occurrence statistics as a language-dependent cue for speech segmentation“, *Dev. Sci.*, let. 20, št. 3, str. e12390-n/a, 2016.
[19] M. Shukla, M. Nespor, in J. Mehler, „An interaction between prosody and statistics in the segmentation of fluent speech.“, *Cogn. Psychol.*, let. 54, št. 1, str. 1–32, feb. 2007.
[20] S. Benavides-varela in J. Mehler, „Verbal Positional Memory in 7-Month-Olds“, *Child Dev.*, let. 86, št. 1, str. 209–223, 2015.
[21] A. D. Endress, M. Nespor, in J. Mehler, „Perceptual and memory constraints on language acquisition.“, *Trends Cogn. Sci.*, let. 13, št. 8, str. 348–53, avg. 2009.
[22] A. D. Endress, B. J. Scholl, in J. Mehler, „The role of salience in the extraction of algebraic rules.“, *J. Exp. Psychol. Gen.*, let. 134, št. 3, str. 406–19, avg. 2005.
[23] A. M. Kovács in J. Mehler, „Cognitive gains in 7-month-old bilingual infants.“, *Proc. Natl. Acad. Sci. U. S. A.*, let. 106, št. 16, str. 6556–60, apr. 2009.
[24] A. M. Kovács in J. Mehler, „Flexible learning of multiple speech structures in bilingual infants“, *Science*, let. 325, št. 5940, str. 611–2, jul. 2009.
[25] A. Karmiloff-smith, „Development itself is the key to understanding developmental disorders“, *Trends Cogn. Sci.*, let. 2, št. 10, str. 389–398, 1998.

[26] U. Goswami, „Sensory theories of developmental dyslexia: three challenges for research“, *Nat Rev Neurosci*, let. 16, št. 1, str. 43–54, jan. 2015.

[27] R. Hari in H. Renvall, „Impaired processing of rapid stimulus sequences in dyslexia.“, *Trends Cogn. Sci.*, let. 5, št. 12, str. 525–532, dec. 2001.

[28] A. D. Baddeley, „Modularity, working memory and language acquisition“, *Second Lang. Res.*, let. 33, št. 3, str. 299–311, 2017.

[29] I. Peretz in M. Coltheart, „Modularity of music processing“, *Nat. Neurosci.*, let. 6, št. 7, str. 688–691, 2015.

[30] J. Zerilli, „Neural redundancy and its relation to neural reuse“, *Philos. Sci.*, let. 86, št. 5, str. 1191–1201, 2019.

[31] M. L. Anderson, „Neural reuse : A fundamental organizational principle of the brain“, *Behav. Brain Sci.*, let. 33, str. 245–313, 2010.

[32] M. L. Anderson in B. L. Finlay, „Allocating structure to function: the strong links between neuroplasticity and natural selection“, *Front. Hum. Neurosci.*, let. 7, št. January, str. 1–16, 2014.

[33] J. Zerilli, „Neural Reuse and the Modularity of Mind : Where to Next for Modularity ?“, *Biol. Theory*, 2018.

[34] A. Guida, G. Campitelli, in F. Gobet, „Becoming an expert: Ontogeny of expertise as an example of neural reuse“, *Behav. Brain Sci.*, let. 39, 2016.

[35] D. S. Grayson *idr.*, „Structural and Functional Rich Club Organization of the Brain in Children and Adults“, *PLoS One*, let. 9, št. 2, str. 1–13, 2014.

[36] R. Siugzdaite, J. Bathelt, J. Holmes, in D. E. Astle, „Transdiagnostic Brain Mapping in Developmental Disorders“, *Curr. Biol.*, str. 1–13, 2020.

[37] T. Donovan, K. Dunn, V. M. Reid, A. Penman, in R. J. Young, „Fetal eye movements in response to a visual stimulus“, *Brain Behav.*, let. 10, št. e01676, str. 1–6, 2020.

[38] V. M. Reid idr., „The Human Fetus Preferentially Engages with Face- like Visual Stimuli Report The Human Fetus Preferentially Engages with Face-like Visual Stimuli“, Curr. Biol., let. 27, št. 12, str. 1825-1828.e3, 2017.

# Vpliv informacije o ceni na subjektivno oceno zvoka violin

Influence of Price Information on the Subjective Evaluation of Violin Sound

Anja Šerbec
Gimnazija Bežigrad
Peričeva ulica 4
Ljubljana, Slovenija
aanjaserbec@gmail.com

## POVZETEK

V raziskavi sem analizirala, v kolikšni meri informacija o ceni inštrumenta vpliva na posameznikovo subjektivno oceno zvoka. Zanimalo me je tudi, ali so subjektivne ocene zvoka pri glasbenikih bolj povezane s ceno violin v primerjavi z ocenami poslušalcev, ki se z glasbo ne ukvarjajo. S poskusom sem preverjala, če bo lažna informacija o ceni vplivala na subjektivno oceno zvoka. Pri poskusu, ko cena ni bila podana, sem zaznala šibko do zmerno povezanost med ceno violine in subjektivno oceno zvoka. Pri poskusu, ko je cena bila podana, sem zaznali visoko povezanost med ceno in subjektivno oceno zvoka.

Posameznikovo vrednotenje zvoka je tako pri glasbenikih kot tudi pri udeležencih, ki se z glasbo ne ukvarjajo močno povezano z informacijo o ceni. Violina, ki sem jo enkrat predstavila z njeno realno prodajno ceno, drugič pa kot bistveno dražjo, je bila drugič ocenjena zaznavno boljše. Najcenejša violina je bila v poskusu, v katerem je bila cena podana, ocenjena zaznavno slabše.

## KLJUČNE BESEDE

placebo efekt, marketing, vplivi na zaznavanje, ocenjevanje violin, informacija o ceni

## ABSTRACT

In this study, I investigated the extent to which an instrument's price information affects a person's attitude toward its sound. I was also interested in whether musicians' ratings of sound aesthetics were more strongly related to violin prices than were the ratings of participants who were not involved with music. I experimented with whether misinformation about price would influence ratings of sound. In the experiment in which price was not mentioned, I found a low to moderate correlation between violin price and sound ratings. In the experiment where price was mentioned, I found a high correlation between price and sound ratings.

Sound ratings correlated strongly with price information for both musicians and non-musicians. The violin we presented once with its actual retail price and a second time as being significantly more expensive was rated significantly better the second time. The cheapest violin was rated significantly worse in the experiment in which the price information was given.


*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*



## KEYWORDS

placebo effect, marketing, effects on sound perception, assessment of violins, price information

## 1 UVOD

Drage stvari so nam pogosto všeč. Mogoče višjo ceno povezujemo z boljšo kakovostjo izdelka, za nekatere pa je posedovanje dragega izdelka statusni simbol. Zdi se, da že sama cena vpliva na naše vrednotenje izdelkov. V raziskavi sem opazovala, kako informacija o ceni vpliva na mnenje poslušalca o zvoku violine. Zanimalo me je, če in v kolikšni meri je poznavanje cene povezano s subjektivno oceno zvoka šestih violin popolnoma različnih cenovnih razredov.

Osnovna predpostavka v ekonomiji je, da je stopnja ugodja pri uživanju nekega produkta odvisna le od lastnosti tega produkta in stanja posameznika. Tako naj bi na primer užitek, ki izhaja iz uživanja pijače bil odvisen le od molekulske sestave pijače in stopnje žeje posameznika [6]. Pretekle raziskave pa so pokazale, da informacije iz okolja vplivajo na naše **pričakovanje in zaznavanje** na senzoričnih področjih: bolečina, vid, vonj in tudi sluh. Kljub temu ni popolnoma znano, kako možgani spremembe pričakovane vrednosti pretvorijo v spremembe izkušene vrednosti [10].

V raziskavi na Stanfordski Univerzi leta 2007 so testirancem povedali, da bodo degustirali pet različnih vin in, da je namen poskusa preučiti vpliv časovnega trajanja degustacije na zaznan okus. Eno izmed vin je bilo degustirano dvakrat: enkrat z realno informacijo o ceni in drugič z (lažno) nizko ceno. Testiranci so bili pozvani, naj poročajo o všečnosti in intenzivnosti okusa vin. Rezultati so pokazali bistvene razlike v oceni všečnosti okusa dveh degustacij istega vina predstavljenega z dvema različnima cenama. Sklepamo, da informacija o ceni znatno vpliva na všečnost okusa. Poskus so izvedli še enkrat, le da so tokrat opazovali delovanje različnih možganskih centrov ob poskušanju vina. Izkazalo se je, da je delovanje možganskih centrov povezanih z sprejemanjem senzoričnih signalov in njihovo predelavo različno pri dveh degustacijah istega vina, ko je informacija o cenah podana [6]. Tudi raziskava z energijskimi pijačami na Stanfordski univerzi iz leta 2005 je predhodno

pokazala, da imajo določene marketinške poteze, kot je določanje in spreminjanje cen vpliv na naše zaznavanje, presojo in vedenje [9]. Pojav je bil poimenovan »marketinški placebo efekt«, saj je zelo podoben znanemu fenomenu placebo efekta v farmaciji [6], [9].

## 2 TEORETIČNE OSNOVE

V prispevku nas, podobno kot v prej opisanih poskusih, zanima fenomen »placebo efekta«, le da se osredotočamo na zaznavanje prijetnosti zvoka.

Placebo efekt je definiran kot »sprememba bolnikovega stanja, ki jo je mogoče pripisati simboličnem vnosu zdravljenja in ne farmakološkim ali fiziološkim lastnostim zdravljenja« [3, pp.1]. Kljub temu, da je pojem placebo efekt ponavadi uporabljen v povezavi z zdravili, je povezan z našim problemom, saj opisuje vpliv informacijskega nabora iz okolja na čutne izkušnje. Opisala bom tudi katere lastnosti zvoka zaznavamo.

### 2.1 Teorija pričakovanja

Teorija pričakovanja pravi, da testirančeva pričakovanja in prepričanost v dober rezultat sprožijo placebo efekt. V skladu s to teorijo bi na primer testiranec iz skupine, ki pozna ceno pričakoval boljši zvok violin, ki so bile predstavljene kot dražje. S prepričanostjo v dober rezultat in pristranskostjo bi jih zato ocenil kot boljše [3].

### 2.2 Klasično pogojevanje

Teorija predvideva, da je placebo efekt pogojni refleks zaradi ponavljajočih se povezav med pogojnim dražljajem (nevtralna komponenta) in brezpogojnim dražljajem (aktivni element) [3]**.** V našem primeru je pogojni dražljaj **informacija o** cen**i** in brezpogojni dražljaj zvok, oziroma **kakovost** zvoka. Testiranci bi lahko **nezavedno prevzeli**, da imajo dražje violine boljši zvok in jih je posledično bolj prijetno poslušati (brezpogojni odziv). Tako bi že sama informacija o višji ceni (pogojni dražljaj) sprožila večjo všečnost do poslušane violine. Seveda velja tudi obratno: če bo imel testiranec negativne izkušnje z cenejšimi violinami, bo nižjo ceno podzavestno povezal z slabšim zvokom.

### 2.3 Socialni vplivi na zaznavanje in vedenje

Pomembno lahko vplivajo na zaznavanje tudi članstvo in procesi v skupini [6]. Kljub temu, da poskus ni bil izveden v skupinah, ampak ga je vsak testiranec reševal sam, menim, da so socialni dejavniki imeli močan vpliv na rezultate. Veliko ljudi je namreč prepričanih, da visoka cena violine kaže, da večina visoko vrednoti to violino. Predvidevam, da bodo namesto, da bi se odločili avtonomno prilagodili mnenje skupini, oziroma temu, kar menijo da je mnenje večine.

### 2.4 Lastnosti zvoka violin

Kljub temu, da se v raziskavi ukvarjam z vplivom informacije o ceni na všečnost zvoka violin in ne sámo kakovost zvoka, ne moremo zanemariti precejšnje verjetnosti, da imajo dražje violine dejansko bolj kvaliteten zvok. Violine se ocenjuje po treh dimenzijah: odzivnost, enakomernost in »glas«. Slednji je izrazito subjektiven, zato vrednosti violin in kakovost zvoka ni mogoče objektivno oceniti [1].

Kljub temu, da se zaznavanje zvoka razlikuje od posameznika do posameznika raziskava na UWE Bristol iz leta 2005 kaže na določeno stopnjo strinjanja pri kvalitativnih opisih lastnosti zvoka inštrumentov pri skupini glasbenikov [4]. Glasbeniki za opis »barve« zvoka (tembre) določene violine pogosto uporabijo »diferencialne pridevnike«. Primeri teh so: svetlost, trdost, jasnost, tankost, polnost, nazalnost, odprtost, ostrina, celo »kovinskost« in »lesenost« zvoka. Glasbenik bi zvok izbrane violine ocenil na dimenzijah: svetel – temen, trd – mehek, jasen – nejasen (»umazan«), tanek – širok, poln – prazen (»na površju«), nazalen – usten, zaprt – odprt. Umestitev zvoka violine na prej-naštetih dimenzijah omogoča glasbenikom bolj poenoteno oceno zvoka izbrane violine v primerjavi z laiki. Uporaba naštetih lasnosti pri ocenjevanju s strani glasbenikov je v raziskavi nakazana pri odgovorih na vprašanje kombiniranega tipa »Kaj je vplivalo na vašo odločitev?«. Na to vprašanje so glasbeniki večkrat odgovorili s pridevniki »čistost«, »mehkoba«, »jasnost«, »odprtost«. Pri posameznikovi oceni pomembno vlogo igrajo osebne preference, a v splošnem velja, da ima dobra violina svetel, mehek, jasen, širok, poln, usten in odprt zvok [1].

## 3 OPIS RAZISKAVE

Kot merski instrument sem uporabila spletni anketni vprašalnik, ki je vseboval poseben tip vprašanja, ki je omogočilo testirancu razvrščanje violin glede na njihovo subjektivno oceno zvoka. Vprašalnik je vseboval tudi zvočni zapis narejen z visoko kakovostnim snemalnikom zvoka Zoom h1. Zvočni zapis je predstavljal posnetke lestvice in melodij, zaigranih na 6 različnih violin (Tabela 1). Vse violine so bile posnete v istem prostoru (predavalnica 212, UL PeF), na njih pa sem igrala z istim lokom.

Vprašalnik je bil sestavljen iz dveh delov: v prvem delu (Poskus 1) so bili podani zgolj posnetki melodij: izseki iz skladb Bacha, Paganinija, Glazunova ter Mozarta. Bach je skladatelj baroka, Glazunov romantike, Mozart klasicizma, Paganini pa sicer spada v romantiko, vendar igranje njegovih Cappricciov ponazarja zmožnost inštrumenta, da se odzove na tehnično zahtevnih delih. Želela sem namreč predstaviti zvok vsake violine v različnih glasbenih slogih. Med glasbeniki namreč velja prepričanje, da nekatere violine bolje »ustrezajo« določenim slogom kot drugim.

Poslušalci so s funkcijo »povleci in spusti« razvrstili šest različnih violin glede na njihovo subjektivno oceno zvoka posamezne violine. Udeleženci so violine med sabo primerjali in jih razvrstili od najboljše do najslabše glede na njihovo oceno estetike zvoka (Slika 1). Povprečne ocene so bile izračunane po naslednjih formulah:

$$nenormirana\_ocena(violina_i) = \frac{\sum_{j=1}^{N}(MAX_{ocena} - ocena_j(violina_i)}{N}$$

$N$ – Število razvrstitev za $violina_i$

$MAX_{ocena}$ – pri Poskus 1 je enaka 6, pri Poskus 2 je enaka 7

$ocena_j(violina_i)$ – razvrstitev violine i na določeno mesto

Ker pa sta bili $MAX_{ocena}$ pri Poskus 1 in Poskus 2 drugačni (pri Poskus 1 je bila $MAX_{ocena}$ 6, ker so testiranci razvrščali 6 posnetkov 6 različnih violin, pri Poskus 2 pa 7, saj so se posnetki

Violine 3 ponovili), je bilo ocene potrebno normirati. Povprečne ocene so bile normirane od 1 do 100 po naslednji formuli:

$$normira\ ocena(violina_i) = \\ = ROUND(\frac{MAX_{nov} - MIN_{nov}}{MAX_{ocena} - MIN_{ocena}} \\ * (nenormirana\_ocena(violina_i) \\ - MAX_{ocena}) + MAX_{nov})$$

$ROUND$ – zaokroženo

$MIN_{nov} = 1$

$MAX_{nov} = 100$

$MIN_{ocena} = 1$

V drugem delu (Poskus 2) so bili poleg posnetkov melodij podani tudi posnetki lestvice a-mol in informacija o ceni. Pri tem so bili enaki posnetki iste violine (VIOLINA 3) podani dvakrat: enkrat z resnično informacijo oceni (3000 evrov) in enkrat z lažno informacijo o ceni (30.000 evrov). Posnetki lestvice so bili dodani zato, da preusmerijo testirančevo pozornost od dejstva, da je v Poskusu 2 navidezno bila predstavljena ena violina več.

Na koncu obeh poskusov so bili testiranci vprašani o tem, kaj je vplivalo na njihovo odločitev. Vprašanje je bilo kombiniranega tipa, nanj pa so lahko odgovorili z več odgovori:

- »jakost zvoka«
- »barva zvoka (tembre)«
- »dinamične razlike«
- »cena«
- »drugo« (odprtega tipa)

Testirance sem razdelila v dve osnovni skupini: glasbeniki in neglasbeniki. Kot glasbeniki so bili označeni vsi, ki so na vprašanje »Kateri stavek vas opisuje?« odgovorili z enim izmed stavkov:

- »Sem profesionalen–i/-a glasben–ik/-ica in igram inštrument – godalo.«
- »Sem profesionalen–i/-a glasben–ik/-ica in ne igram inštrumenta, ki je godalo.«
- »Obiskujem akademijo za glasbo in igram inštrument – godalo.«
- »Obiskujem akademijo za glasbo in igram inštrument, ki ni godalo.«
- »Obiskujem glasbeno šolo in igram inštrument – godalo.«
- »Obiskujem glasbeno šolo in igram inštrument, ki ni godalo.«
- »Končal-a sem osnovno [in srednjo] glasbeno šolo.«

Kot ne-glasbeniki so bili označeni vsi, ki so na vprašanje »Kateri stavek vas opisuje?« odgovorili z enim izmed stavkov:

- »Obiskoval-a sem nekaj let osnovne glasbene šole.«
- »Ljubiteljsko se ukvarjam z glasbo.«
- »Z glasbo se ne ukvarjam.«

Zanimala so me naslednja raziskovalna vprašanja:

*Vprašanje 1: Ali se zaznavanje estetike zvoka glede na informiranost o ceni pri obeh skupinah (glasbeniki, ne-glasbeniki) razlikuje?*

*Vprašanje 2: Ali so subjektivne ocene zvoka pri skupini glasbenikov v poskusu brez informacije o ceni bolj povezane s ceno violin v primerjavi z ocenami estetike zvoka v skupini ne-glasbenikov?*

*Vprašanje 3: Ali bo napačna informacija o ceni violine (lažna informacija, da je cenejša violina draga) vplivala na subjektivno oceno zvoka pri tako glasbenikih kot tudi ne-glasbenikih?*

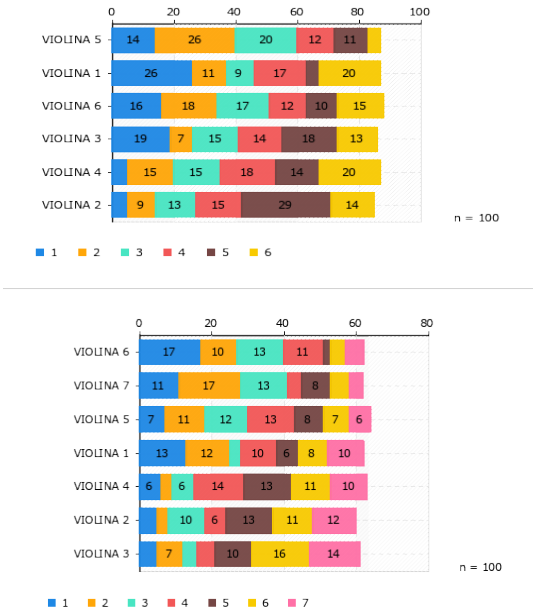
Da bi odgovorila na Vprašanje 1 sem primerjala rezultate Poskusa 1 in Poskusa 2. Odgovor na Vprašanje 2 sem iskala v rezultatih Poskusa 1. Pri odgovarjanju na Vprašanje 3 sem uporabila rezultate Poskusa 2.

**Tabela 1: Maloprodajne cene violin**

| | *Cena (EUR)* |
|---|---|
| *VIOLINA 1* | 16 500 |
| *VIOLINA 2* | 7 200 |
| *VIOLINA 3* | 3 000 |
| *VIOLINA 4* | 13 500 |
| *VIOLINA 5* | 15 200 |
| *VIOLINA 6* | 20 000 |
| *VIOLINA 7 (VIOLINA 3)* | 30 000 (3 000) |

## 3.1 Opis vzorca

Poskus je v večini potekal preko spleta, delno pa tudi v živo na Gimnaziji Bežigrad in Akademiji za glasbo. Vprašalnik je do konca izpolnilo 100 ljudi, od tega 40 glasbenikov in 60 ne-glasbenikov. Reševan je bil v Sloveniji, Makedoniji, Rusiji, Nemčiji in Avstriji. Anketni vprašalnik je v celoti rešilo 4o žensk in 31 moških. Anketni vprašalnik je bil objavljen na neuradni Facebook strani dijakov in bivših dijakov Gimnazije Bežigrad, rešili pa so ga tudi dijaki Konzervatorija za glasbo in balet Ljubljana, študenti in profesorji Akademije za glasbo Ljubljana in Univerze za umetnost Gradec ter člani simfoničnega orchestra RTV Slovenija.

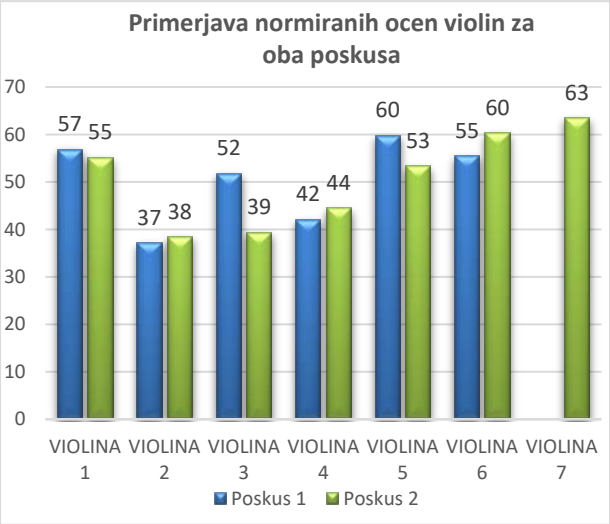



**Slika 1: Razvrščanje violin po okusu od 1 do 6 (7). Zgornja slika kaže frekvence, oz. kako so udeleženci razvrščali violine brez informacije o ceni. Spodnja slika kaže frekvence, oz. kako so udeleženci razvrščali violine z informacijo o ceni.**

## 4 REZULTATI IN UGOTOVITVE

Vpršanje 1: *Ali se subjektivne ocene zvoka violin glede na informiranost o ceni pri obeh skupinah (glasbeniki, ne-glasbeniki) razlikujeje?*

Graf, ki ga prikazuje Slika 1 prikazuje povprečne normirane ocene violin na lestvici od 1 do 100, ki so izračunane na podlagi ocen violin v celotnem vzorcu (torej glasbeniki in ne-glasbeniki).




**Slika 2: Povprečne ocene violin vseh testirancev za oba poskusa (z in brez informacije o ceni**

Da bi ugotovila, če se ocene violin, ki so jih dali testiranci pred in po informiranju o ceni (torej rezultati Poskusa 1 in Poskusa 2) statistično značilno razlikujejo, sem uporabila Wilcoxonov test predznačenih rangov. Ta je pokazal statistično značilno razliko med rezultati Poskusa 1 in Poskusa 2 pri violinah 1, 3, 4, 5 ($p < 0,05$). Test ni pokazal statistično značilne razlike med rezultati Poskusa 1 in Poskusa 2 pri Violini 2 in Violini 6 ($p > 0,05$). Teh izjem ne morem pojasniti.

Rezultati Wilcoxonovega testa predznačenih rangov nakazujejo, da se ocene večine violin glede na informiranost o ceni v celotnem vzorcu razlikujejo.

**Tabela 2: Wilcoxonov test predznačenih rangov za pare ocen violin, ki so jih dali testiranci pred in po informiranju o ceni. Stat. pomembne vrednosti so označene krepko.**

| ***Wilcoxonov test*** | *Brez informacije o ceni/ Z informacijo o ceni* |
|---|---|
| *VIOLINA 1* | **$z$ = 2,119 $p$ = 0, 034** |
| *VIOLINA 2* | $z$ = 1,678 $p$ = 0,092 |
| ***VIOLINA 3*** | **$z$ = 3,910 $p$ = 0,0001** |
| *VIOLINA 4* | **$z$ = 2,208 $p$ = 0,027** |
| *VIOLINA 5* | **$z$ = 2,951 $p$ = 0,003** |
| *VIOLINA 6* | $z$ = 0,528 $p$ = 0,597 |

Uporabila sem Spearmanov koeficient korelacije za oceno povezanosti med ceno violin in oceno zvoka violin v celotnem vzorcu. V Poskusu 1 ni bilo statistično značilne korelacije med spremenljivkama cena violin in ocena zvoka violin, $r_s$ = 0,564; $p$ = 0,188; N = 6. Korelacija med spremenljivkama je bila v Poskusu 1 zmerna. V Poskusu 2 sem zaznala statistično značilno korelacijo med spremenljivkama cena in subjektivna ocena zvoka, $r_s$ = 0,964; $p$ = 0,0004; N = 6. Korelacija med spremenljivkama je bila v Poskusu 2 zelo močna.

To indicira, da je bila ocena zvoka v celotnem vzorcu pri Poskusu 2 povezana z informacijo o ceni violin.

Spearmanov koeficient korelacije za oceno povezanosti med ceno violin in oceno zvoka violin sem izračunala za vsako skupino posebaj. Pri skupini glasbenikov pri Poskusu 1 ni bilo statistično pomembne korelacije med spremenljivkama cena in subjektivna ocena zvoka, $r_s$ = 0,771; $p$ = 0,072; N = 6. Korelacija med spremenljivkama je bila v Poskusu 1 močna. V Poskusu 2 sem pri skupini glasbenikov zaznala statistično pomembno korelacijo med spremenljivkama cena in subjektivna ocena zvoka, $r_s$ = 0,886; $p$ = 0,019; N = 6. Korelacija med spremenljivkama je bila v Poskusu 2 zelo močna.

Pri skupini neglasbenikov pri Poskusu 1 ni bilo statistično pomembne korelacije med spremenljivkama cena in subjektivna ocena zvoka $r_s$ = 0,314; $p$ = 0,544; N = 6. Korelacija med spremenljivkama je bila v Poskusu 1 šibka. V Poskusu 2 sem pri skupini glasbenikov zaznala statistično pomembno korelacijo med spremenljivkama cena in subjektivna ocena zvoka, $r_s$ = 0,943; $p$ = 0,005; N = 6. Korelacija med spremenljivkama je bila v Poskusu 2 zelo močna.

To indicira, da je bila ocena zvoka v vsaki od skupin pri Poskusu 2 povezana z informacijo o ceni violin.

**Tabela 3: Spearmanov koeficient korelacije za oceno povezanosti med ceno violin in oceno zvoka violin pri skupinah glasbenikov in neglasbenikov. Statistično pomembne korelacije so označene krepko.**

| ***Spearmanov koeficient za 6 violin*** | *Brez Informacije o ceni* | *Z informacijo o ceni* |
|---|---|---|
| *Glasbeniki* | $r_s$ = 0,771, $p$ = 0,072 | **$r_s$ = 0,886, $p$ = 0,019** |
| *Neglasbeniki* | $r_s$ = 0,314, $p$ = 0,544 | **$r_s$ = 0,943, $p$ = 0,005** |

Na vprašanje »Kaj je vplivalo na vašo odločitev?« pri Poskusu 2 so testiranci lahko odgovorili z več odgovori. Prikazan delež testirancev je izbral naslednje odgovore:

- »jakost zvoka« - 26,15%
- »barva zvoka (tembre) – lestvica« - 44,25%
- »barva zvoka (tembre) – melodije« - 49,28%
- »dinamične razlike« - 24,14%
- »cena« 15,8%
- »drugo« (odprtega tipa) -17,10%

Dejavnik, ki je po mnenju testirancev najbolj vplival na njihovo razvrstitev je bila barva zvoka (tembre) pri posnetkih melodij (43 odgovorov). Veliko vlogo naj bi igrala tudi barva zvoka (tembre) pri lestvicah (38 odgovorov). Pod »drugo« so bili pogosti odgovori: »alikvoti«, »izenačenost registrov«, »odzivnost« ter »intonacija«. Zanimivo je, da je cena med

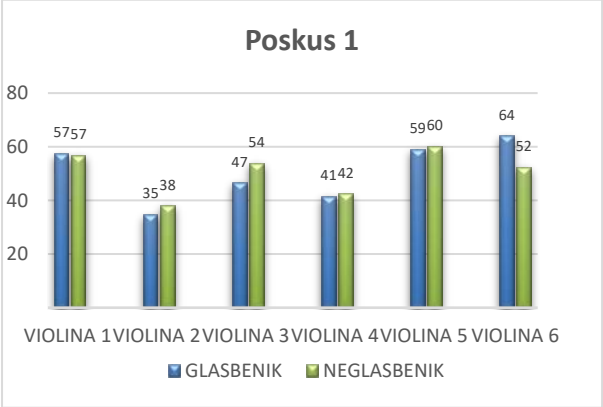
dejavniki, ki po mnenju testirancev vplivajo na njihovo razvrstitev, po pogostosti na zadnjem mestu z le 14 odgovori..

Vprašanje 2*: Ali so subjektivne ocene zvoka violin pri skupini glasbenikov v poskusu brez informacije o ceni bolj povezane s ceno violin v primerjavi z ocenami estetike zvoka v skupini ne-glasbenikov?*

Izračunan Spearmanov koeficient korelacije med ceno violin in oceno zvoka violin pri skupini glasbenikov je pri Poskusu 1 kazal močno korelacijo (Tabela 3).

Spearmanov koeficient korelacije med ceno violin in oceno zvoka violin je pri skupini neglasbenikov pri Poskusu 1 kazal zgolj zmerno korelacijo med spremenljivkama (Tabela 3).

Ker pa noben od omenjenih koeficientov ni statistično značilen, ne morem poročati o povezanosti med spremenljivkama pri obeh skupinah.




**Slika 3: Primerjava povprečnih normiranih ocen violin skupin glasbeniki in ne-glasbeniki pri Poskusu 1**

Vprašanje 3: *Ali bo napačna informacija o ceni violine (lažna informacija, da je cenejša violina draga) vplivala na oceno zvoka pri tako glasbenikih kot tudi ne-glasbenikih*?

Da bi ugotovila, če je razlika v ocenah zvoka pri celotnem vzorcu pred in po informiranju o ceni (enkrat z resnično informacijo oceni in enkrat z lažno) statistično značilna sem uporabila Wilcoxonov test test predznačenih rangov.

Wilcoxonov test predznačenih rangov je nakazoval na statistično značilno razliko v ocenah zvoka pri celotnem vzorcu pred in po informiranju o ceni, ko je bila podana resnična informacija o ceni. Ocena zvoka po informiranju o resnični ceni je bila zaznavno nižja, $z = 3{,}886$; $p = 0{,}0001$.

Presenetljivo pa je, da Wilcoxonov test predznačenih rangov ni indiciral statistično značine razlike v ocenah zvoka pri celotnem vzorcu pred in po informiranju o ceni, ko je bila podana lažna informacija o ceni (zlagano visoka). Ocena zvoka po informiranju o lažni ceni ni bila zaznavno višja, z = 0,247; p = 0,802.

**Tabela 3: Wilcoxonov test za VIOLINO 3/7 celoten vzorec. Stat. pomembne vrednosti so označene krepko.**

| ***Wilcoxonov test*** | *Brez informacije o ceni/ Z informacijo o ceni: 3000* | *Brez informacije o ceni/ Z lažno informacijo o ceni: 30.000* |
|---|---|---|
| *VIOLINA 3/7* | ***z* = 3,886**<br>***p* = 0,0001** | *z* = 0,247<br>*p* = 0,802 |

Wilcoxonov test predznačenih rangov za pare ocen violin, ki so jih dali testiranci pred in po informiranju o ceni sem izračunala tudi za vsako skupino posebej.

Wilcoxonov test predznačenih rangov ni indiciral statistično značilne razlike v ocenah zvoka pri skupini glasbenikov pred in po informiranju o ceni, ko je bila podana resnična informacija o ceni. Ocena zvoka po informiranju o resnični ceni ni bila zaznavno nižja, $z = 0{,}809$; $p = 0{,}381$.

Zanimivo je, da je Wilcoxonov test predznačenih rangov indiciral statistično značilno razliko v ocenah zvoka pri skupini glasbenikov pred in po informiranju o ceni, ko je bila podana lažna informacija o ceni. Ocena zvoka po informiranju o resnični ceni je bila zaznavno višja, $z = 2{,}505$; $p = 0{,}012$.

Wilcoxonov test predznačenih rangov je indiciral statistično značilno razliko v ocenah zvoka pri skupini neglasbenikov pred in po informiranju o ceni, ko je bila podana resnična informacija o ceni. Ocena zvoka po informiranju o resnični ceni je bila zaznavno nižja, $z = 4{,}139$; $p = 0{,}000003$.

Presenetljivo je tudi, da Wilcoxonov test predznačenih rangov **ni** indiciral statistično značilne razlike v ocenah zvoka pri skupini neglasbenikov pred in po informiranju o ceni, ko je bila podana resnična informacija o ceni. Ocena zvoka po informiranju o resnični ceni ni bila zaznavno nižja, $z = 1{,}267$; $p = 0{,}205$.

Rezultati nakazujejo, da je lažna informacija o ceni bolj vplivala na glasbenike v primerjavi z neglasbeniki.

**Tabela 4: Wilcoxonov test za VIOLINO 3/7 za glasbenike in neglasbenike. Stat. pomembne vrednosti so označene krepko.**

| ***Wilcoxonov test za 3/7 violino*** | *Brez informacije o ceni / Z informacijo o ceni 3.000 EUR* | *Brez informacije o ceni / Z lažno informacijo o ceni 30.000 EUR* |
|---|---|---|
| *Glasbeniki* | *z* = 0,809<br>*p* = 0,381 | ***z* = 2,505**<br>***p* = 0,012** |
| *Neglasbeniki* | ***z* = 4,139**<br>***p* = 0,000003** | *z* = 1,267<br>*p* = 0,205 |

## 5 MOŽNE IZBOLJŠAVE

Dejstvo, da je bil anketni vprašalnik večinoma reševan preko spleta in ne v živo pa ima nekaj pomanjkljivosti. Testiranci so pri poslušanju zvočnih posnetkov violin imeli različno kakovostno opremo (zvočniki). Testiranci z boljšo opremo so tako lahko bolj natančno slišali razlike v lastnostih zvoka med violinami. Nekaj

pomanjkljivosti pa je bilo tudi v pripravi samega vprašalnika: izpolnjevanje vprašalnika je zaradi dolžine posnetkov vzelo vsaj 12 minut. Posledično del testirancev ni rešil vprašalnika v celoti, kar je močno zmanjšalo obseg vzorca. Možna posledica je tudi to, da je udeleženim proti koncu poskusa zmanjkovalo pozornosti (in potrpljenja) in so zato violine ocenjevali naključno ali po informaciji o ceni. Razlog za daljše posnetke je bila želja, da pri vsaki violini predstavim njen zven v različnih stilih preko melodij iz različnih obdobji glasbene umetnosti.

Kot moteča spremenljivka, bi lahko deloval tudi vpliv izvajalca: ker sem bila sama izvajalka, nisem bila enako »navajena« na vse igrane violine. Nekatere violine so bile redno servisirane, strune na njih so bile nove in bile so »igrane«, druge pa ne. Vsi našteti faktorji zaznavno vplivajo na kakovost zvoka violine.

Da bi poskus izboljšala, bi ga izvedla še enkrat, z nekaj spremembami: vse violine bi servisirala in »uigrala«. Da izničim vpliv lastne afinitete do določenih violin, bi tokrat posnela igranje violinista, ki na vse violine igra prvič. Uporabila bi bolj kakovosten snemalnik zvoka. Poskus bi najraje izvedla v živo in tako zagotovila, da vsi udeleženci poslušajo posnetke pod enakimi pogoji (enako kakovostne slušalke/zvočnik). Zanimivo bi bilo tudi razširiti poskus na področje nevro-ergonomije in z slikanjem možganov z metodo funkcijske magnetne resonance (fMRI) opazovati razlike v delovanju možganov testirancev pri poslušanju violin in odločanju.

Ob ponovnem izvajanju poskusa bi v anketni vprašalnik vključili več vprašanj o lastnostih testirancev. Tako bi vzorec razdelili na več smiselnih podskupin, ki bi jih primerjali med seboj. (Npr. “Na testirance, mlajše od 25 let, je informacija o ceni vplivala bolj/manj, kot na testirance starejše od 25 let.”)

Znano je, da je ocena kakovosti zvoka inštrumenta zelo kompleksna tema: pri njej igrajo vlogo barva, jakost, dinamične razlike, idr. Veliko vlogo igrajo tudi osebne preference, zato je določanje vrednosti violine nekakšna »siva cona«. V poskusu sem opazovala vpliv faktorja, ki ni neposredno povezan z lastnostmi zvoka: informacija o ceni. Raziskava zato omogoča nekoliko provokativen pogled v svet prodaje in kupovanja violin, ter je uporabna tako za izdelovalce in prodajalce kot za kupce violin.

Uporabna je tudi na področju psihologije v marketingu, saj nakazuje, da informacije iz okolja vplivajo na naša pričakovanja povezana z vrednostjo in na to kako poročamo o izkušnjah na senzoričnih področjih, natančneje na področju sluha. Predvidevam, da bi spoznanja raziskave lahko prenesli še na druga senzorična področja, kot so okus, vid, vonj.

## REFERENCE


[1] Bissinger, G. (2008). Structural acoustics of good and bad violins. *The Journal of the Acoustical Society of America*, *124*(3), 1764-1773.

[2] Brody, H. (1981). The Lie That Heals: The Ethics of Giving Palcebos. *Soc. Resp.: Journalism L. Med.*, *7*, 27.

[3] Cho, H. J., Hotopf, M., & Wessely, S. (2005). The placebo response in the treatment of chronic fatigue syndrome: a systematic review and meta-analysis. *Psychosomatic Medicine*, *67*(2), 301–313.

[4] Darke, G. (2005, March). Assessment of timbre using verbal attributes. In *Conference on Interdisciplinary Musicology. Montreal, Quebec*. sn.

[5] Karmarkar, U. R., & Plassmann, H. (2019). Consumer neuroscience: Past, present, and future. *Organizational Research Methods*, *22*(1), 174.

[6] Kompare, A., Stražišar, M., Dogša, I., Vec, T., & Curk, J. (2019). *Psihologija: spoznanja in dileme: učbenik za psihologijo v 4. letniku gimnazijskega izobraževanja*. DZS.

[7] Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, *105*(3), 1050–1054.

[8] Schacter, D., Gilbert, D., Wegner, D., & Hood, B. M. (2011). *Psychology: European Edition*. Macmillan International Higher Education.

[9] Shiv, B., Carmon, Z., & Ariely, D. (2005). Placebo effects of marketing actions: Consumers may get what they pay for. *Journal of marketing Research*, *42*(4), 383–393

[10] Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, *27*(3), 161–168

# AI Art: Merely a Possibility or Already a Reality?

Tadej Todorović
Faculty of Arts
University of Maribor
Maribor, Slovenia
tadej.todorovic@gmail.com

Janez Bregant
Faculty of Arts
University of Maribor
Maribor, Slovenia
janez.bregant@um.com

## ABSTRACT

The paper discusses the compatibility of AI art with various definitions of art within the analytic tradition, namely functional, historical, and institutional ones. For every definition, we first offer a general overview, discuss whether AI art could be compatible with it, detect possible problems, and finally offer real-life examples that could arguably serve as an example of AI that fits the given definition. In the final section, we address the issue of intentionality for AI art, which seems to be in one way or another part of all discussed definitions and which seems to be the biggest challenge for AI art.

## KEYWORDS

Artificial intelligence, art, functionalism, historicism, institutionalism, intentionalism.

## 1 Introduction

Today, there are hardly any doubts that artificial intelligence (AI) can perform many tasks much better than us, all the way from playing chess, backgammon, or checkers to intelligent scheduling and pricing systems in airline reservations, proving theorems, or solving equations. And as the AIs are getting better and better at these domain-specific tasks, we, with more and more uncertainty, diligently move the goalposts, stating that AI will surely not be able to beat us at the next mark. No wonder then that one of the last bastions of human uniqueness, i.e. creativity, best shown through art and its creations, is fiercely defended against the possibility of AI art. What should philosophy say about that? Are there any definitional obstacles to admitting AI art? Are there already existing examples of AI art that might fit various definitions of art?

Definitions of art remain a controversial subject in analytic philosophy. There has been much discussion about the value of the definition of art and many sceptical concerns about its existence in the first place, starting all the way back in the 1950s [1]. Nevertheless, the AI art debate is a debate about whether AI can produce art, so it has to presuppose that there are in fact works of art and that there is an intelligible way or definition that can capture this phenomenon. Not presupposing this would render the entire debate meaningless.

However, to remain as metaphysically non-committing as possible, we decided to analyse the compatibility of AI art with various most popular definitions. We excluded some more basic definitions, namely single property definitions, such as representational, expressive, and formal definitions; these seem to have fallen out of fashion, undoubtedly because they are "not difficult to find fault with" [1].

Thus, we first analyse the compatibility of AI art with functional definitions, followed by historical and institutional definitions of art, and offering existing AI art examples along the way. Afterwards, we also offer a response to probably the biggest obstacle to AI art, i.e. intentionality.

## 2 AI and functional definitions of art

Functional definitions of art define art in terms of some function or intended function. Usually, the function is connected with some aesthetic properties, such as the aesthetic experience we undergo when admiring a work of art, e.g., catharsis or simply some aesthetic judgments or experiences. In this sense, functional definitions are more traditional, and have issues accommodating, e.g., modern art, like Duchamp's ready-mades (although some have argued that ready-mades have aesthetic properties [2]). Despite their flaws, such definitions seem to be perfect for accommodating AI art. Beardsley's definition can serve as a good example of a functional aesthetic definition. It states that an artwork is "either an arrangement of conditions intended to be capable of affording an experience with marked aesthetic character or (incidentally) an arrangement belonging to a class or type of arrangements that is typically intended to have this capacity" [3].

But which conditions evoke such feelings and experiences? To our knowledge, a satisfactorily account of them has not been given. Nevertheless, in the context of AI art, there seem to be no formal obstacles against AI creating (art) works that meet such conditions. In fact, this is not only conceivable, but has arguably already been done. A prime example is the "Creative Adversarial Network" (CAN) [4], the design of which was motivated by Berlyne's theory [5] inspired by his most significant arousal-



*Information Society 2020, 5–9 October 2020, Ljubljana, Slovenia*

raising properties for aesthetics: "novelty, surprisingness, complexity, ambiguity, and puzzlingness".

The CAN project proved very successful. The authors ran a series of experiments (Turing style tests with human subjects) with the created artworks to test how the AI measures up to human artists. The experiment III is the most relevant for functional definitions. In it, they asked human subjects to rate the artworks by CAN and artworks by human artists (set of paintings from a display at Art Basel 2016). The paintings were rated on a scale of 1-5 (5 being the best) on intentionality, visual structure, communication, and inspiration. Not only did the human subjects fail to notice that CAN paintings were not made by human artists, they outperformed human artists in all metrics. Of course, the size of the experiment was rather small (21 participants), so the results are not statistically powerful; however, as the authors state, "the fact that subjects found the images generated by the machine intentional, visually structured, communicative, and inspiring, with similar levels to actual human art, indicates that subjects see these images as art!" [4]

Functional definitions do not require anything but the realization of certain functional, i.e. aesthetic, properties, which makes them tailor-made for AI art. We believe it is safe to claim that if one subscribes to such a definition, they would be hard-pressed to find an argument against including the already-existing AI artworks.

## 3 AI and historical definitions of art

Historical definitions are another popular way of understanding art. The core message of historical definitions is that an artwork "is standing in some specified art-historical relation to some specified earlier artworks" [1], which is similar to family-resemblance theories in certain aspects [6]. Moreover, and this is what distinguishes historical definitions from functional or institutional definitions: proponents of historical definitions do not commit to a trans-historical concept of art, i.e. the concept that would capture commonalities across various classes of artworks in distinct historical periods, e.g. some stable core of aesthetic properties that are present in all art movements throughout the history. Thus, historical definitions present "an alternative to the definitional approach" [7]. One of the most recognised historical definition of art is offered by Levinson, who defines a work of art as "something that has been intended by someone for regard or treatment in some overall way that some earlier or pre-existing artwork or artworks are or were correctly regarded or treated" [8].

There seem to be two common elements in historical definitions (even though proponents of historical definitions understand their reasoning as an alternative to the definitional approach, we will refer to historical "definitions" as definitions for the sake of simplicity and because our argument does not hinge on this): let us call the first one the family-resemblance element, and the second one the intentional element, despite the fact that some historical definitions do not require the intentional element [9].

In this section, we will focus on the family-resemblance element; however, we will address intentionality as a problem for AI art in the final section. The question that we have to answer is thus whether AI artworks could stand in an appropriate relationship to established artworks and, more importantly, whether they already do. Similar to the problem in functional definitions, this should not present an insurmountable problem for AI art. It is not only conceivable that AIs could use a family-resemblance process to create artworks, AIs already utilize a process that looks extremely similar. Alexander Mordvintsev, the software engineer behind DeepDream, Google's neural network, writes as follows, "We train an artificial neural network by showing it millions of training examples and gradually adjusting the network parameters until it gives the classifications we want" [10].

The already mentioned CAN is an even better example: it uses a slightly different approach because its purpose is to create artworks that would be indistinguishable from human artworks. The CAN is comprised of two adversary networks, a discriminator and a generator. A discriminator is "trained" on human art samples, so it has a reference of art images, accompanied with styles and labels. The generator then creates new works of art, trying to accomplish two things: the first is to generate works that the discriminator would recognize as works of art, i.e., it tries to create art that fits into the already-existing styles. However, if it did only that, it would just emulate artworks, similar to an art forger. So, the second task of the CAN generator is to confuse the discriminator regarding the style of the work created. So, "on one hand it tries to fool the discriminator to think it is 'art', and on the other hand it tries to confuse the discriminator about the style of the work generated" [4]. In other words, the neural network has to navigate between the Scylla, which is getting recognized as art, and Charybdis, which is generating works that are "style-ambiguous", trying to find the sweet spot where the painting still resembles other works of art but it is still original. And considering the experimental results introduced in the previous section, CAN is apparently doing an extremely good job at it.

The idea of AI art being compatible with the historical definitions is thus not only conceivable or possible; just like with functional definitions of art, we could reasonably state that there are already examples of AI art that fit the criteria of historical definitions.

## 4 AI and institutional definitions of art

The institutional definition of art is probably one of the most influential and simultaneously one of the most criticized definitions of art of the 20th century. Many have argued that "the definition's obvious circularity is vicious" [1]; nevertheless, it has remained fairly popular. The groundwork for institutionalism was laid by Danto [11]; however, Dickie's institutional definition is probably the most influential. The spirit of institutionalism can be summed up by the following quote: "a work of art is an artifact which has had conferred upon it the status of candidate for appreciation by the artworld" [12]. In other words, something is a work of art if people within the artworld grant it such a status. The definition is more elaborate, and has been expanded by Dickie in his more recent work, so it now consists of five

interlocking conditions: "(1) An artist is a person who participates with understanding in the making of a work of art. (2) A work of art is an artifact of a kind created to be presented to an artworld public. (3) A public is a set of persons the members of which are prepared in some degree to understand an object which is presented to them. (4) The artworld is the totality of all artworld systems. (5) An artworld system is a framework for the presentation of a work of art by an artist to an artworld public" [12].

For brevity's sake, we will only focus on the premises that seem problematic for AI art, i.e., premises (1) and (2). Premise (1) seems problematic, as AI is obviously not a person. However, the context has to be considered here; the authors of the 20th century assumed that "the artist is always human, without exploring much whether non-humans can create art" [13]. This seems fairly anthropocentric in this day and age, and we are confident that most theorists would agree that a being with the same or greater understanding in the making of a work of art (e.g., aliens) would still be considered artists. Therefore, the problem does not seem to be not being human, but rather not possessing the capacity to understand and partake in the making of a work of art. This is closely (if not completely) related to intentionality, which we address in the next section, so we will put it aside for now.

The second premise might also pose some problems. There seem to be two separate questions here: what counts as an artifact and is an AI made object an artifact. So what is an artifact? Hilpien's definition should serve our goals: "artifacts are physical objects which have been manufactured for a certain purpose or intentionally modified for a certain purpose" [14]. Notice that such a definition "does not rule out the possibility that at least some things made by non-human animals are artifacts" [15]. E.g. "[b]eavers /…/ might be thought to intentionally construct dams in order to create ponds" [15]. On the other hand, some more rigid behaviours of other animals, like webs woven by spiders, might not count as artifacts. Paths can serve as an even more ambiguous example. They are often created unintentionally, when people take the same short-cut across the university lawn over and over again: but, as Preston argues, "/…/ what is the point of saying that such a path is not an artifact, whereas an identical one that was created intentionally by exactly the same process is? Moreover, what would it take to make the erstwhile non-artifactual path into an artifact? Would it be enough to notice and approve it? Or would I have to intentionally maintain it, by sweeping it clean of leaves, for instance?" [15] The line has to be drawn somewhere, and it is hard to imagine that the line will not be, in some sense, arbitrary.

So, are AI made objects artifacts? If we dismiss the artifacts debate because it seems arbitrary, then it does not matter. If one insists on the artifact/non-artifact distinction, a proponent of such distinction has to first offer a good reason in favour of it. Even if such a reason could be provided, they have to answer the following question: how to classify AI object that are indistinguishable from human artifacts? If someone not familiar with The Painting Fool [16] discovered a painting made by it, they would, without a doubt, classify it as a (human) artifact. So why should we revoke that status once we discover that there was no intention involved in the production of the image? It would be almost as difficult as arguing that the path that was created unintentionally somehow differs as an artifact from the intentionally created path. In short, if humans recognize something as an artifact and behave as if it is an artifact, then why should we not count it as one? The idea that something is an artifact if recognized as an artifact is also compatible with Dickie's institutionalism since, according to him, "anything brought into an art space as a candidate for appreciation becomes thereby 'artefactualized'" [17].

The only question that remains to answer is whether there are examples of AI art that pass fit the institutional definition. And, in fact, there are. Jeff Clune decided to test the level of artworks produced by Evolving Artificial Intelligence Lab's deep neural networks (DNN), submitting the artworks to the University of Wyoming's 40th Annual Juried Student Exhibition, "which accepted 35.5% of its submissions" [18]. Its artworks were not only accepted, but also among the "21.3% of submissions to receive an award" [18], and, what is perhaps most important for an institutional definition of art, were displayed at the university's art museum. So not only can we say that there does not seem to be a good reason against AI art in the framework of the institutional definition of art, we could arguably claim that AI art is already here.

## 5 AI and intentionality

Some sort of intentionality component was present in almost all analysed definitions. The idea that something can only count as art if it was produced intentionally could thus be compatible with all analysed definitions. Intentionality is aboutness, it is "power of minds and mental states to be about, to represent, or to stand for, things, properties and states of affairs" [19]. It is hard to imagine that an organism or a system would possess such powers without consciousness. Even consciousness is not sufficient for intentionality: we agree that animals (most animals) are conscious, but they (or babies) do not possess intentionality, as intentionality belongs to higher order cognition. So, we cannot possibly ascribe intentionality to AI, as we have no reason to think it is even conscious.

Nevertheless, we believe intentionality is problematic as a condition for artworks. Here's why. Definitions of art usually include intentionality to exclude natural phenomena being art. However, intentionality can be understood in two ways. We can understand it in the narrower sense of producing and expressing a particular idea that the artist has, or we can understand it in a much broader, abstract sense of simply creating a work of art. If one stick to the former, this already excludes many art movements. Surrealism greatly emphasized automatism, which is "perhaps the most famous of their [surrealists's] techniques for evading conscious control of the artistic process" [20]. Breton defines Surrealism as "Psychic automatism in its pure state, by which one proposes to express /…/ the actual functioning of thought /…/ in the absence of any control exercised by reason, exempt from any aesthetic or moral concern" [21]. So not only did the surrealists want to create artworks in the absence of reason and intention, they saw "reason as a guard barring entry to this storehouse" [20].

Defenders of intentionality can quickly offer the following retort: even if one admits that surrealists' process for creating art was not intentional in the narrow sense, they nevertheless had intention to produce art in the broader sense; they had the more abstract "impulse" or "urge" to create a work of art, which AI lacks. However, not all artists throughout the history demanded or valued such broad intentionalism; in fact, some have been explicitly against it. Advocates of Primitivism, a widespread trend in the modern art, celebrated "primitive works", which "came from an unconscious source of creativity rather than from artistic traditions, an idea which suited many modern artists /…/ modern artists also praised the 'primitivism' of art produced by children, the insane and untrained, 'naïve' adults." [20]. Even though one could argue that "primitive" art had a source of inspiration, a sort of intentionality, it would be hard to argue that what they had in mind was this broader concept of creating art. Such a broader claim would be even harder to defend in case of children or the "insane".

Two conclusions can be drawn from all this: if one demands intention in the narrow sense then this would exclude movements like Surrealism, and therefore should not be a necessary condition for artworks; and if one demands intentionality in the broader sense then such a concept will differ massively over cultures and individuals, especially if we find value in "primitive" art. If modern artists cherished and valued art produced by children and the "insane", which lack intentionality in the broader sense altogether, then this should also not be a necessary condition for artworks.

Throughout this paper, we have shown examples of AI artworks that were not only appreciated as art, but which also won prizes, and arguably outperformed human artists. Spectators recognized such works as intentional, inspiring, and communicative. Similar to "primitive" art, AI was able to achieve this without intentionality in the narrow or broader sense. Understanding intentionality in the narrow sense excludes too much from the world of art, and understanding it in the broader sense does not allow an objective definition of art: the concept of this artistic impulse, as seen with Primitivism, just varies too much across cultures and individuals to enable an unbiased description of art. As such, it would seem more appropriate to judge works of art on their external properties, not the intentions of the artists.

We can confidently say that AI (art) works can already pass some kind of the so-called Turing test in the world of art, something that perhaps many post-modern or contemporary human works of art would not. And whereas some people see AI art as blasphemous, we see it as potentially offering us new insight into our understanding of art. Nevertheless, it seems that whatever objection AI defeats, the goal-post always moves further away. Simon Colton wrote (about his creation, the Painting Fool) that "it is our hope that one-day people will have to admit that the Painting Fool is creative because they can no longer think of a good reason why it is not" [16]. Similarly, hopefully one-day people will have to admit that AI can produce art, because they can no longer think of a good reason why it could not.

## REFERENCES


[1] Adajian, Thomas. 2018. "The Definition of Art". In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2018. https://plato.stanford.edu/archives/fall2018/entries/art-definition/.

[2] Lind, Richard. 1992. "The Aesthetic Essence of Art". *Journal of Aesthetics and Art Criticism*, 50, pp. 117–29. DOI: https://doi.org/10.1017/CBO9780511625251.023.

[3] Beardsley, Monroe. 1982. *The Aesthetic Point of View*. Ithaca, New York: Cornell University Press.

[4] Elgammal, Ahmed, Liu, Bingchen, Elhoseiny, Mohamed, Mazzone, Marian. 2017. "CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms". *International Conference on Computational Creativity (ICCC)*.

[5] Berlyne, Daniel Ellis. 1971. *Aesthetics and psychobiology*. Appleton-Century-Crofts. DOI: 10.1093/oxfordhb/9780195304787.001.0001.

[6] Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. G.E.M. Anscombe and R. Rhees (Eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.

[7] Carroll, Noel. 1993. "Historical Narratives and the Philosophy of Art". *The Journal of Aesthetics and Art Criticism*, 51(3), pp. 313–26.

[8] Levinson. 2007. "Artworks as Artifacts.". In E. Margolis and S. Laurence (Eds.), *Creations of the Mind: Theories of Artifacts and Their Representation*, Oxford: Oxford University Press, pp. 74–82.

[9] Stock, Kathleen and Thomson-Jones, Katherine. 2008. *New Waves in Aesthetics*. London: Palgrave Macmillan.

[10] Mordvintsev, Alexander. 2015. "Inceptionism: Going Deeper into Neural Networks". *Google AI Blog*. Available at: https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[11] Danto, Arthur. 1981. *The Transfiguration of the Commonplace*. Cambridge: Harvard University Press.

[12] Dickie, George. 1974. *Art and the Aesthetic: An Institutional Analysis*. Ithaca, NY: Cornell University Press.

[13] Hertzmann, Aaron. 2018. "Can Computers Create Art?". *Arts* 7 (2): 18. DOI: https://doi.org/10.3390/arts7020018.

[14] Hilpinen, Risto. 1992. "Artifacts and Works of Art". *Theoria*, 58(1), pp. 58–82. DOI: 10.1111/j.1755-2567.1992.tb01155.x.

[15] Preston, Beth. 2020. "Artifact". In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2020. URL = <https://plato.stanford.edu/archives/fall2020/entries/artifact/>.

[16] Colton Simon. 2012. "The Painting Fool: Stories from Building an Automated Painter". In John McCormack and Mark d'Inverno (Eds.) *Computers and Creativity*. Berlin: Springer. DOI: https://doi.org/10.1007/978-3-642-31727-9_1

[17] Slater, Barry Hartley. "Aesthetics". In James Fieser and Bradley Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Available at: https://iep.utm.edu/aestheti/.

[18] Shein, Esther. 2017. "Computing the Arts". *Communications of the ACM*, 60(4), pp. 17–19. DOI: 10.1145/3048381.

[19] Jacob, Pierre. 2019. "Intentionality". In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2019. URL = <https://plato.stanford.edu/archives/win2019/entries/intentionality/>.

[20] Little, Stephen. 2004. *Isms: Understanding art*. New York: Universe.

[21] Breton, André. 1969. *Manifestoes of Surrealism*. Ann Arbor: The University of Michigan Press

# Compliance with COVID-19 preventive behaviors and proneness to cognitive biases

Manca Toporišič Gašperšič[†]
Undergraduate Psychology Program
Faculty of Arts, University of Ljubljana (Slovenia)
manca.verano@gmail.com

Nataša Grof
Poljane grammar school
Strossmayerjeva 1
Ljubljana, Slovenia
natasa.grof@gimnazija-poljane.com

## ABSTRACT

Due to common non-compliance with behavioral hygiene recommendations to contain the SARS-CoV-2 virus, the younger generation has often been regarded as a catalyst of the current pandemic. Therefore, the aim of the present study is to determine the connection between proneness to specific cognitive biases and compliance with COVID-19 preventive recommendations in high school students. Our results indicate that decision myopia is positively correlated to non-compliance with COVID-19 containment measures. Surprisingly, no link has been found between risk aversion and compliance to self-protective recommendations, whilst individuals who are more prone to belief bias report greater compliance with COVID-19 preventive behaviors. The results clearly indicate that proneness to cognitive biases is somewhat important but not a decisive factor of adherence to preventive measures.

## KEYWORDS

COVID-19, preventive behavioral measures, compliance, cognitive biases, high school students

## 1 INTRODUCTION

### 1.1 Theoretical background

**Known psychological correlates to compliance with behavioral interventions**

With the rise of novel coronavirus SARS-CoV-2 variants and related vaccine hesitancy trends, basic behavioral hygienic measures (such as wearing masks, frequent hand washing, as well as physical distancing) have remained the fundamental tools to contain the spread of the virus. However, evidently certain individuals do not comply to these behavioral recommendations [1], thus probably contributing to the spread of the coronavirus. Identifying factors that are linked to compliance with behavioral recommendations and restrictions is thus extremely important.



*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*


Research so far has extensively focused on linking certain personal traits to compliance with behavioral recommendations. Extraversion has therefore been negatively correlated to compliance with COVID-19 social distancing measures, whereas conscientiousness is believed to be positively correlated to compliance [2]. At the same time, low levels of empathy and antisocial traits are linked to noncompliance with containment measures [3, 4]. On the other hand, current literature has offered inadequate understanding of the cognitive factors of behavioral non-compliance. In this study, we try to theoretically and empirically bridge this research gap. We therefore undertake to examine certain cognitive biases we believe might be related to engaging in self-protective behavior.

**Cognitive biases and their possible correlation with preventive behavior**

Framing is defined in the framework of prospect theory, which predicts that people are inconsistent when evaluating losses and gains. In particular, when faced with losses, people typically tend to engage in more risk-seeking behavior than when faced with gains [5, 6]. In consequence, more negative, loss-emphasizing information may result in greater risk-taking decision making. Adherence to even the most basic hygienic measures which aim to limit the spread of the coronavirus SARS-CoV-2, is to a certain extent a decision based on one's risk attitude. In the current pandemic, the most recurrent example of framing losses is enumerating the number of lives lost due to COVID-19. Emphasizing saved lives, is on the other hand, an example of framing gains.

However, in addition to this typical framing context, some authors have already pointed out other framing types. There have been indications that different countries framed the outbreak differently at the beginning of the coronavirus outbreak in 2020. Whilst Western countries focused more on framing COVID-19 as a respiratory disease, similar to the seasonal flu, Asian countries compared the novel coronavirus to the SARS virus – a difference in framing that supposedly contributed to the great success of Asian countries in flattening the initial curves of new infections [7].

Risk aversion is another important notion, defined in the framework of prospect theory. It is a cognitive bias, best described as a constant inclination to select the most certain and reliable option, even when there are more profitable (but at the same time riskier) options available [8]. Current theory stipulates that people more prone to this bias, tend to be more compliant with COVID-19 measures [9].

On the other hand, not engaging in self-protective behavior may not only be connected to one's risk attitudes, but also to their lack of reasoning and unwillingness to incorporate new evidence into their thought processes. In syllogistic reasoning, belief bias is described as the tendency "to rely on prior beliefs rather than to fully obey logical principles [10]." In other words, it means being constrained by your own opinions and predispositions. In response, acquiring new, accurate, and unbiased information can be extremely difficult for individuals who are especially prone to this cognitive bias [11]. And since the COVID-19 pandemic has often been referred to as a pandemic of misinformation [12], possessing factual evidence that may be connected with our health-related decisions is surely of utmost importance.

In contrast, decision myopia or the present bias "is the nonlinear and inconstant tendency of many individuals to prefer a smaller sooner pay-off over a larger future pay-off [13]." Favoring smaller and sooner rewards over long-term ones has been a recurrent phenomenon of the pandemic. During the pandemic, we have witnessed how many people have disproportionately ignored social distancing guidelines in order to socially interact with others. However, since social gatherings are known to lead to a spike in coronavirus cases, this a very short-sighted move on various levels since it is believed to additionally contribute to lives lost. In addition, long-lasting draconian lockdowns to contain the spread of the virus limiting in-person contacts are often imposed to restrict such gatherings.

### 1.2 Overall aim and hypotheses

The key objective of the study is to shed light on the relationship between framing, belief bias, risk aversion, and decision myopia to non-compliance with behavioral recommendations[1] to contain the spread of the coronavirus. According to the presented theory, we introduce several hypotheses. On account of framing effects and their role in risky decision making, we hypothesize:

**H1a:** Participants who are exposed to the framing of losses, will make riskier choices than participants who are exposed to the framing of gains in the neutral condition.

**H1b:** Participants who are exposed to the framing of losses when seasonal flu is mentioned, will make riskier choices than participants who are exposed to the framing of gains when virus SARS is mentioned.

**H1c:** Participants who are exposed to the framing of gains in the neutral condition, are less likely to opt for the riskier option than participants who are exposed to the framing of gains when the SARS virus is mentioned.

**H1d:** Participants who are exposed to the framing of losses in the neutral condition, are less likely to opt for the riskier option than participants who are exposed to the framing of losses when the SARS virus is mentioned.

Moreover, our other hypotheses are as follows:

**H2:** Participants, prone to belief bias, report lower compliance with COVID-19 containment behavioral recommendations.

**H3:** There is a positive correlation between compliance with behavioral guidelines and loss aversion.

**H4:** Decision myopia is negatively correlated with compliance with behavioral recommendations to contain the spread of COVID-19.

## 2 METHODOLOGY

### 2.1 Participants and procedure

To determine the cognitive factors of non-compliance with behavioral guidelines in the younger generation, the generation often proclaimed to be reluctant towards the epidemiological restrictions [1], our study exclusively focused on this age group. The study thus included 83 participants – all students at Poljane Grammar School, aged from 15 to 19 years old. However, as three participants failed to complete the study, their results were excluded from the final analysis. The majority (75%) of participants identified themselves as female, 24% defined themselves as male, whilst the remaining 1% did not wish to disclose their gender. Although this gender structure is not typical of the general population, it is typical of Poljane Grammar School.

The empirical study was conducted on 18th and 19th February 2021 via the Slovenian survey tool *1ka*. Since the study took place during the national COVID-19 lockdown and in-person learning restrictions, the subjects completed the study in the course of their class meetings that were held online, and were a part of their distance-learning schedule. All participants were informed about and consented to the general purpose of the study, and were acquainted with the fact that their participation in the research was entirely voluntary and anonymous.

While completing the empirical questionnaire, they were supervised by the researcher via Zoom, the online video conferencing platform used by their high school. Whilst the research was being carried out, all participants were required to turn on their camera. Moreover, all the participants were notified that any communication among them was prohibited since it could adversely affect the results. To prevent interpersonal communication among the participants, we carefully set the Zoom chat settings so that they prevented participants from communicating with each other. At the same time, a direct online chat communication channel between each participant and the researcher was established. Thus, students participating in the study were able to point out certain technical issues or other concerns directly to the researcher without disrupting others. Furthermore, students were not externally motivated in any way

[1] In this paper, we distinguish between *basic behavioral recommendations* to contain the spread of the coronavirus (for instance, hand washing, mask wearing, and maintaining physical distance from others) and *restrictive measures* (such as lockdown, curfews, and regional restrictions). In our study, we overall address non-compliance to basic behavioral recommendations, but not non-compliance to restrictive measures. Partially our decision is based on the fact that restrictive measures are of limited use when individuals are non-compliant with the basic behavioral recommendations. A study [27] has, for instance, indicated that basic behavioral recommendations can be epidemiologically as successful as restrictive containment measures, provided that individuals adhere to these recommendations, we add. On the other hand, our decision to focus on behavioral interventions rather than on restrictive measures was also largely based on the fact that an international extension of the current study will probably be carried out. As epidemiological (restrictive) measures vary from country to country, a goal of the present study was also to lay out the measurements for our later studies.

to participate in the study: they were not given a fee nor were their results classified or publicly disclosed in any way. We therefore believe that the current results are the best possible representation of the participants' proneness to cognitive biases.

## 2.2 Tasks and measures

**Compliance with COVID-19 behavioral recommendations**

To measure reported compliance with the COVID-19 behavioral containment recommendations, we used an adapted form of the *Compliance with COVID-19 prevention guidelines scale* [14]. The adapted 4-point Likert scale includes 13 items, which predominantly focus on determining the extent of compliance with basic hygiene guidelines (such as mask wearing or hand washing) rather than on compliance with more restrictive measures (for instance curfews or lockdown).

**Framing**

In the framing section of the questionnaire, participants were randomly assigned into two groups. We measured the impact of framing with two similar tasks. The first task was the original task used by Kahneman and Tversky [6]. In this paper, we often refer to this task of framing as framing in the *neutral condition*. The instructions of the task were identical in both experimental groups and are, as follows:

Imagine that Slovenia[2] is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed.

Participants of both tasks were then asked to peruse scientific estimates of how many people would die / live if a certain program is accepted and make a decision on which program should be imposed. In both experimental groups, programs actually predict the same number of lives lost / lives saved. However, as indicated below, gains (lives saved) were framed in the 1st experimental group, whilst losses (lives lost) were framed in the 2nd experimental group. That is:

**Group 1:** If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

**Group 2:** If Program A is adopted, 400 people will die.

If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

In addition to the task in the neutral condition, another task was added to measure how specifying the disease impacts risk-seeking behavior. The text of the second task was slightly modified in comparison with the first task. Participants in the experimental group 1 (the group with framing of gains) were given the information that the disease of the outbreak is similar to diseases, caused by the SARS virus. In contrast, participants in the group with framing of losses were provided with the comparison of the disease with the seasonal flu.

**Belief Bias**

To measure a participant's proneness to belief bias and its connection to compliance with behavioral recommendations, we used adapted tasks of Markovits and Nantel [15]. Although the original toolkit to measure this cognitive bias was comprised of eight tasks, we used only seven of them as we believed that participants would generally not be acquainted with the individual mentioned in one task[3], and hence unable to respond to the question. All seven questions used were in fact syllogisms – combinations of three statements. The participants were instructed to assume that the first two statements (premises) are true; their task was to estimate whether or not the third statement is the right conclusion derived from the first two statements.

In four tasks, the conclusion that is correctly derived from the two premises is contradictory to general knowledge. As such, proneness to belief bias is in these tasks determined as the willingness to estimate conclusions as inaccurate due to their dissimilarity to generalized facts. This can be illustrated by the following task used in the study:

Premise 1: All things that are smoked are good for your health.

Premise 2: Cigarettes are smoked.

Conclusion: Cigarettes are good for the health.

If we were to ignore the premises and read only the conclusion, we would correctly proclaim it to be false. However, the conclusion is in accordance with the premises, hence it is correct in the context of the given task. A person, susceptible to belief bias will, consequently, likely struggle to reflect on the intuitively-suggested responses and in the particular case incorrectly answer that the conclusion is false.

On the other hand, the other three tasks we used had seemingly reasonable conclusions. However, these conclusions could not have been made on the basis of the given premises and were, as a result, incorrect. Here, proneness to belief bias is regarded as the decision that the conclusion is right. This can be exemplified by the following task:

Premise 1: All flowers have petals.

Premise 2: Roses have petals.

Conclusion: Roses are flowers.

**Risk aversion**

We used a truncated Holt-Laury Task [16][4] to measure risk aversion. The task is formulated as a set of paired lottery choices and was initially designed to measure financial risk aversion. However, it is applicable to non-financial fields as well, and as such useful for the purpose of our study, as people are consistent in their preferences regarding risk-taking in all areas of life [17].

The original task contains ten rounds of paired choices, whilst ours included only nine due to the complexity and length of the study. In every round, participants are required to opt for either option A or option B; both options are profitable. Nevertheless, their profitability and risk level differ. The potential profits of both options remain constant throughout all nine rounds (thus,

[2] The original text of the task predicted that the U.S., and not Slovenia was preparing for an outbreak. For the purpose of this study, this detail was changed.

[3] This individual was John D. Rockefeller.

[4] In comparison with the original task, the currency was also changed to familiarize the participants with the task.

option A can potentially bring either €2.00 or €1.60, whereas the predicted payoff of option B is €3.85 or €0.10; [16]). Generally, option A is regarded as the "safe" option, meanwhile option B is regarded as the "riskier" option as the potential profits in option B vary more than potential profits in option A [8]. The course of the task can be demonstrated by its first three rounds:

**Table 1: First three rounds of Holt-Laury Task**

| Round | Option A | | Option B | |
|---|---|---|---|---|
| 1 | 10% chance of receiving €2.00 | 90% chance of receiving €1.60 | 10% chance of receiving €3.85 | 90% chance of receiving €0.10 |
| 2 | 20% chance of receiving €2.00 | 80% chance of receiving €1.60 | 20% chance of receiving €3.85 | 80% chance of receiving €0.10 |
| 3 | 30% chance of receiving €2.00 | 70% chance of receiving €1.60 | 30% chance of receiving €3.85 | 70% chance of receiving €0.10 |

The average behavior of the majority of participants in initial rounds is to opt for the safer option, option A. This trend is, however, expected to alter when the likelihood of receiving larger payments as a result of choosing option B substantially increases [16]. One's willingness to engage in risk-taking behavior is measured by the number of "risky" decisions – the selections of option B.

Decision myopia

An adapted [5] measurement of intertemporal choice by Frederick [18] was used in this study in order to link decision myopia to non-compliance with behavioral recommendations to contain the spread of SARS-CoV-2. In total the measurement included eleven items. As with the risk aversion task, the intertemporal choice measurement was structured as a combination of paired lottery choices.

In the first eight rounds, participants had to choose between two profitable options, option A and option B. Option B was always more lucrative than option A. However, the payoff of option A was always immediate or at least chronologically sooner in comparison with the payoff of option B. For instance:

If you choose option A, you will receive €3000 this month. If you choose option B, you will receive €3400 next month.

In such tasks, short-sighted individuals are therefore expected to select instant gratification by persistently choosing option A [18]. In the 9th and 10th round, participants were asked to choose between the two given options once again. This time both options were loss-making: option A predicted a more immediate, but financially lower loss, whilst option B involved a greater, but deferred loss. Decision myopic individuals are believed to prefer deferred losses even when it is not financially profitable for them [18], as indicated in the following example:

If you choose option A, you will lose €1000 this year. If you choose option B, you will lose €2000 next year.

The 11th round was in fact not a lottery choice task – it was a question, also used in the original intertemporal choice measurement [18], which asked the participants to indicate whether they would be prepared to pay more for overnight shipping of a chosen product.

## 3 RESULTS

All acquired data were statistically analyzed in Microsoft Excel 2016.

### 3.1 Framing

To measure the impact of framing gains/losses, we used a chi-squared test. Our data indicate that there is no statistical difference in risk taking behavior when losses are framed as opposed to gains in the neutral condition, $X^2$ (1, N = 80) = 0.03, p = 0.87. Moreover, no significant difference in risk attitude has been found when comparing the framing of gains when SARS is mentioned and the framing of losses when seasonal flu is mentioned, $X^2$ (1, N = 80) = 0.00, p = 0.99.

Nevertheless, the results demonstrate that participants who were exposed to framing of gains in the neutral condition were more risk averse than participants who were exposed to framing of gains when the SARS virus was mentioned, $X^2$ (1, N = 80) = 26.53, p = 0.00. On the other hand, the difference in risk attitudes is statistically significant when comparing framing of losses in the neutral condition to framing of losses when the seasonal flu was mentioned; when the flu is mentioned, participants tend to acquire select the risk-taking option more commonly, $X^2$ (1, N =80) = 4.82, p = 0.03.

### 3.2 Cognitive biases and compliance

Correlations between belief bias, loss aversion, decision myopia, and compliance with COVID-19 preventive recommendations are measured with the Pearson correlation coefficient. Data analysis showed that proneness to belief bias and compliance with behavioral recommendations are positively correlated (r = 0.35, p < 0,01). However, there is no statistically significant correlation between proneness to risk aversion and compliance (r = 0.09, p = 0.42). Furthermore, a negative correlation has been found between decision myopia and compliance to COVID-19 preventive behavioral recommendations (r =-0.53, p < 0.01).

## 4 DISCUSSION

Our study has offered a more profound understanding of behavior during the ongoing pandemic. To provide an accurate insight, we exclusively focused on the correlation between proneness to certain cognitive factors and compliance with preventive measures. However, we acknowledge the fact that our

[5] In addition to the fact that the task was shortened (original task to measure intertemporal choice included 17 items), we also changed the currency – as with the risk aversion task.

results might have been affected by other equally important factors correlated to compliance, such as demographic characteristics, socioeconomic status, personality, individual differences in the perception of and emotional responses to the pandemic, resilience, political ideology, conspiracy mentality etc. Furthermore, our study has also shed light on some results that differ from those in the current literature.

Our results, for instance, did not confirm that in the neutral condition, participants exposed to framing of losses were, in consequence, more in favor of engaging in risk-taking behavior than their peers exposed to framing of gains. This is contrary to the pre-existing theory [5, 6]. Similarly, no significant results were found when comparing framing of losses and framing of gains with regard to seasonal flu and the SARS virus. We were thus not able to confirm our first two hypotheses. In our opinion, there are several possible reasons for such results. Firstly, participants in our study were high school students, who are not often represented in gain-loss framing research. It is therefore possible that the impact such framing has on high school students is limited. At the same time, we must acknowledge that the students, representatives of the younger generation, were perhaps not so familiar with the SARS virus, which might impact their uptake of risky / safe options. Secondly, the experiment took place during the ongoing COVID-19 pandemic. It is possible that participants were either very disturbed by reading the outbreak scenario (which might have been, to a certain extent, reminiscent of the current pandemic) or indifferent towards it, as people may become when unable to help others in need [19].

In contrast, it is very interesting that participants' risk attitudes noticeably change when a specific disease is mentioned. Analyzed data indicate that specifying the disease as very similar to either the SARS virus or the seasonal flu contributes to subjects engaging in risk-taking decision making, no matter whether losses or gains are framed. Since our study included only high school students, we cannot transpose these findings to the general population. However, it seems that in the risk-loss framework our subjects understood every specification of the disease as a loss, which caused them to engage in more risk-taking behavior.

Our results are unanticipated in terms of other hypotheses as well. Contrary to our initial expectations, individuals who are more prone to belief bias express greater level of compliance with COVID-19 behavioral recommendations. This might be linked to the fact that, during the pandemic, compliance with COVID-19 behavioral recommendations and regulations has, in many cases, become a political matter [20]. Previous research has shown that people who overall tend to reflect less on their decisions (a characteristic of belief bias) often support populistic leaders [21]. In that regard, the decision to comply with behavioral recommendations and containment measures might be more politically motivated than health-related. This is additionally confirmed by the pre-existing literature in social psychology: individuals who support the group imposing the conformity, are more likely to conform to their social norms as well [22]. Furthermore, a handful of studies in the field report similar results - participants who are less reflective in their decision making (that is, they rely more on their intuition than on analytical deliberation when making decisions) are reportedly more compliant with preventive measures [23, 24]. Such results remain unaccounted for: it is not clear whether they can be directly linked to the use of heuristics, mental shortcuts, as simply complying rather than questioning the measures often requires less cognitive effort, or there is an indirect correlation between cognitive reflection and proneness to biases, compliance, and other noteworthy psychological factors, such as social norms [24].

Similarly unexpected was the finding that students prone to risk aversion bias were not more inclined to comply with behavioral recommendations. The current literature on preventive behaviors suggests that the perceived threat that COVID-19 presents to an individual is a significant factor of compliance to preventive measures [25]. In other words, when feeling threatened, people typically engage in more risk-averse behavior than when they feel there is no danger. According to the national tracking data of the spread of coronavirus SARS-CoV-2 in Slovenia, COVID-19 presents a relatively low threat to the population of high school students [26]. This may, in turn, impact their risk attitudes and compliance with preventive measures. At the same time, it is important to stress that the measuring tool used to estimate the extent of participants' risk aversion was designed to measure financial risk attitudes. Although inclination towards risk-taking behavior has been found to be consistent in every behavioral aspect [17], there is a possibility that we would have obtained significant results, if we had introduced a measuring tool for health-related risk attitudes. This is certainly an important fact we need to consider before planning our future research in the field.

Our finding that impulsive satisfaction of needs is linked to non-compliance with COVID-19 preventive measures is in line with the current literature. It has been suggested that the proneness to this cognitive bias should be used to promote stay-at-home restrictions and recommendations by providing free internet access or benefit packages for vulnerable groups [13].

Overall, our study offers an intriguing and thought-provoking insight into cognitive correlates of COVID-19 preventive behaviors and is a valuable starting point for future research in the field.

## REFERENCES


[1] Syon Bhanot. 2020. Why are people ignoring expert warnings? - Psychological Reactance. *Behavioral Scientist*. https://behavioralscientist.org/why-are-people-ignoring-expert-warnings

[2] Lucas de F. Carvalho, Giselle Pianowski, and André P. Gonçalves. 2020. Personality differences and COVID-19: are extroversion and conscientiousness personality traits associated with engagement with containment measures?. *Trends Psychiatry Psychother*, Vol. 42(2), 179-184. DOI: https://doi.org/10.1590/2237-6089-2020-0029

[3] Fabiano K. et al. 2021. Compliance with containment measures to the COVID-19 pandemic over time: Do antisocial traits matter? *Personality and individual differences*, Vol. 168, 110346. DOI: https://doi.org/10.1016/j.paid.2020.110346

[4] Marcin Zajenkowski et al. 2020. Who complies with the restrictions to reduce the spread of COVID-19?: Personality and perceptions of the COVID-19 situation. *Personality and individual differences*, Vol. 166, 110199. DOI: https://doi.org/10.1016/j.paid.2020.110199

[5] Daneil Kahneman. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux. New York: NY.

[6] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, Volume 47(2), 263-291.

[7] Olivier Sibony. 2020. *HEC Paris Insights - Cognitive Biases and Decision-Making in the COVID-19 Crisis*. [Video]. https://www.youtube.com/watch?v=RhAKzmpwpzw

[8] Urška Ferjančič. 2019. *The influence of hormones and personal traits on the propensity for risk-taking* (Masters thesis, University of Ljubljana,

School of Economics and Business). http://www.cek.ef.uni-lj.si/magister/ferjancic3490-B.pdf

[9] Mahmoud A. H. Shahin, and Rasha M. Hussien. 2020. Risk perception regarding the COVID-19 outbreak among the general population: a comparative Middle East survey. *Middle East Curr Psychiatry*, Vol. 27(71). DOI: https://doi.org/10.1186/s43045-020-00080-7

[10] Ding Daoqun et al. 2020. Belief Bias Effect in Older Adults: Roles of Working Memory and Need for Cognition. *Front. Psychol*. 10:2940. DOI: https://doi.org/10.3389/fpsyg.2019.02940

[11] Laura Macchi et al. 2019. How to Get Rid of the Belief Bias: Boosting Analytical Thinking via Pragmatics. *Europe's Journal of Psychology*, Vol. 15(3), 595–613. DOI: https://doi.org/10.5964/ejop.v15i3.1794

[12] Fabio Tagliabue, Luca Galassi, and Pierpaolo Mariani. 2020. The "Pandemic" of Disinformation in COVID-19. *SN comprehensive clinical medicine*, 1–3. Advance online publication. DOI: https://doi.org/10.1007/s42399-020-00439-1

[13] Moslem Soofi, Farid Najafi, and Behzad Karami-Matin. 2020. Using Insights from Behavioral Economics to Mitigate the Spread of COVID-19. *Applied health economics and health policy*, Vol. 18(3), 345-350. DOI: https://doi.org/10.1007/s40258-020-00595-4.

[14] Nejc Plohl and Bojan Musil. 2020. Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. *Psychology, Health & Medicine*, 1-12. DOI: https://doi.org/10.1080/13548506.2020.1772988.

[15] Henry Markovits and Guilaine Nantel. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, Vol. 17(1), 11–17. DOI: https://doi.org/10.3758/BF03199552

[16] Charles A. Holt and Susan K. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review*, Vol. 92(5), 1644–1655.

[17] Eyal Ert and Ernan Haruvy. 2017. Revisiting risk aversion: Can risk preferences change with experience? *Economics Letters*, Vol. 151, 91–95.

[18] Shane Frederick. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, Vol. 19 (4), 25-42. DOI: https://doi.org/10.1257/089533005775196732

[19] Daniel Västfjäll et al. 2014. Compassion fade: affect and charity are greatest for a single child in need. *PloS one*, Vol. 9(6), e100115. DOI: https://doi.org/10.1371/journal.pone.0100115

[20] Jordan Steffen and Jiuqing Cheng 2021. The influence of gain-loss framing and its interaction with political ideology on social distancing and mask wearing compliance during the COVID-19 pandemic. *Current psychology* (New Brunswick, N.J.), 1–11. Advance online publication. DOI: https://doi.org/10.1007/s12144-021-02148-x

[21] Gordon Pennycook and David G. Rand. 2019. Cognitive Reflection and the 2016 US Presidential Election. *Personality and Social Psychology Bulletin*, Vol. 45(2) 224–239. DOI: https://doi.org/10.1177/0146167218783192

[22] Deborah J. Terry and Michael A. Hogg. 1996. Group Norms and the Attitude-Behavior Relationship: A Role for Group Identification. *Personality and Social Psychology Bulletin*, *22*(8), 776–793. https://doi.org/10.1177/0146167296228002

[23] Predrag Teovanović et al. 2020. Irrational beliefs differentially predict adherence to guidelines and pseudoscientific practices during the COVID-19 pandemic. *Applied Cognitive Psychology*, *35*(2), 486-496. https://doi.org/10.1002/acp.3770.

[24] Volker Thoma, et al. 2021. Cognitive Predictors of Precautionary Behavior During the COVID-19 Pandemic. Front. Psychol. 12:589800. DOI: https://doi.org/10.3389/fpsyg.2021.589800

[25] Eric Bonetto et al. 2021. Basic human values during the COVID-19 outbreak, perceived threat and their relationships with compliance with movement restrictions and social distancing. *PloS one*, *16*(6), e0253430. https://doi.org/10.1371/journal.pone.0253430.

[26] COVID-19 sledilnik [COVID-19 Tracker Slovenia]. 2021. https://covid-19.sledilnik.org/en/stats.

[27] Nils Haug et al. 2020. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav*, Vol. 4, 1303–1312. DOI: https://doi.org/10.1038/s41562-020-01009-0

# The ecological rationality of probabilistic learning rules in unreliable circumstances

Borut Trpin†
Munich Center for Mathematical Philosophy
Ludwig Maximilian University of Munich
Munich Germany
borut.trpin@guest.arnes.si

Ana Marija Plementaš
MEi:CogSci
University of Ljubljana
Ljubljana Slovenia
ap0231@student.uni-lj.si

## ABSTRACT

In today's flood of information in many fields we do not know which sources are reliable and which are not. On what basis can we draw conclusions? Whom to trust? We could say each of us has a belief system that updates based on the arrival of new relevant evidence. In our research we used a computer model where we were investigating which learning rule is more reliable when we do not have a trustworthy source. The main goal is to discover the truth and to do so quickly. Our results show that different probabilistic learning rules may be preferable in different situations and environments.

## KEYWORDS

ecological rationality, belief updating, reasoning, learning rules, uncertainty

## 1 INTRODUCTION

We cannot fully rely on our senses nor on other external sources of information (e.g., testimony provided by others). In addition, it seems that there are multiple types of reasoning under uncertainty in the sense that we use different learning (or reasoning) rules that guide the process of reasoning. For instance, in trying to reach a conclusion about some question (e.g., a doctor is trying to diagnose a patient) on the basis of some evidence/information (e.g., diagnostic tests) an agent might follow a principle of inferring to the best explanation (e.g., of the tests and their sensitivity and specificity). Another agent might consider other aspects of the situation and hence follow different learning rules like, e.g., how confirmatory the evidence is of some hypothesis that is being reasoned about (e.g., if a patient had a disease X, how likely it would be that the tests would be such and such given the objectively known information about the reliability of the test). A question that may be raised could then be put as follows: Given that there are multiple ways of reasoning under uncertainty (represented by different learning rules as described above), are some types of reasoning under uncertainty better than others and how may we even tell whether one type is better than another? That is, how can we compare the performance of various learning rules that guide our reasoning?

It is quite clear that in answering this question we need to consider what the goals of reasoning are. To name a few possibilities: perhaps the goal of reasoning is to increase the understanding of the phenomenon that is the subject of reasoning, or the goal may be to uncover whether some statement holds. In fact, it seems that there are countless aspects that could be considered as valuable outcomes of reasoning and that could as such be used in comparing which rule that guides reasoning is better (or better in some context).

In our investigation we focused on two valuable outcomes: (i) uncovering the truth, and (ii) the speed of reasoning. The former, (i), considers how certain one is of true propositions due to reasoning according to a specific (learning) rule. If (i) is our guide, then we take a rule to be better if it makes one more certain of true propositions. The latter, (ii), considers how quickly one can reach conclusions while reasoning. Similarly, if a rule is quicker in making an agent more certain (it quickly lessens uncertainty), then it performs better on this count.

Ideally, both (i) and (ii) would go hand in hand: a reasoner would reach true conclusions and would also reach them quickly. However, it seems that they do not usually go hand in hand: rules that are especially conducive of (i) seem to typically not be so conducive of (ii), and vice-versa (see, e.g., [1], [2]): more conservative learning rules (i.e., not jumping to conclusions too quickly) are usually such that lead to more accurate conclusions.

For instance, one could excel on count (i) but fail on count (ii): e.g., a learning rule could lead to mostly true conclusions but only after a vast amount of evidence is considered. An example of this would be a medical doctor that identifies the correct disease in her patient but needs to conduct a large number of diagnostic tests before she is able to do so. Similarly, one could underperform on (i) but excel on (ii): such a case would be a doctor that makes a diagnosis on the basis of a single or few tests but her diagnosis is wrong. What we aimed to answer in our research project was which learning (or reasoning) rules are the most conducive of (i) and (ii), and how the two valuable goals (truth and speed) could be balanced when we compare different learning rules.



*Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia*

Additionally, we wanted to keep in mind that the sources of information need not be fully reliable -- they could even be completely misleading.

## 2 BACKGROUND

Our research project actually starts from an investigation of learning on the basis of partial lying, i.e., learning in cases where one asserts information that she believes to be likely but not necessarily false. Another agent then learns based on both (a) such statements and (b) observations of whether the statements are true or false. This provides the basis for estimating the reliability of the source: if statements are (mostly) true, the source is more reliable and vice versa, if statements are mostly false, the source is taken to be more unreliable. We were interested in reliability/trustworthiness of the source under uncertainty more generally (e.g., diagnostic tests in a medical setting might be unreliable too, not just our interlocutors who may want to mislead us intentionally), but a previous research project of partial lying (see [3]) turned out to be a good starting point because it provided a useful formal description of the mechanisms on how to estimate reliability/trustworthiness of a source and how to incorporate this estimate in a learning rule (there: Bayesian learning, although our research project also includes other learning rules). Before we can explain why partial lying is very similar to the topic we were investigating, let us briefly explain the issue of what partial lying even is.

Philosophers define lying with four conditions: (1) a statement, (2) the belief that the statement is false, (3) the addressee, and (4) the purpose of misleading the addressee (see [3] and the references therein).

If someone is constantly lying to us, this individual can be simply deemed unreliable and ignored or even taken as if they are telling us the opposite of truth (saying "A" could be taken as evidence for "not A"). If, however, truth and lies are mingled in varying proportions, choosing whether to trust this individual, and if so to what degree, becomes increasingly difficult. This fact has been emphasized by Trpin and colleagues [1], which pointed out that the definition for lying misses out on many similar cases because the second condition is too strict. They broadened the second condition – we usually also consider someone a liar when they believe their statement to be more likely false than true. However, as they discovered through several computer simulations, estimating the trustworthiness of the source then becomes more difficult, hence such medium-strong lies (that is, those where the liar is only somewhat certain that they are asserting falsehoods) do us more epistemic harm. Following Bayes' learning rule to model lying, the research conducted by Trpin and colleagues [1] sparked debates as to whether it is sensible to consider partial lies at all, if one aims to reduce epistemic harm. What they found is that this approach is only useful when the goal is to quickly avoid believing false propositions.

As we can see from this brief explanation of the background, they already considered both valuable outcomes of reasoning that we were also interested in: (i) epistemic aspects: how close to truth we get due to learning/reasoning under uncertainty, and (ii) pragmatic aspects: how quickly we manage to form strong beliefs on the basis of learning. Moreover, they considered unreliable sources similarly as we did.

Another related research project was conducted by Douven (see [4]). In that research, the focus was not so much on unreliable sources of information but rather on how different probabilistic learning rules compare. In the first part of his research, which was based on computer simulations of learning, he found that the rules diverged on both aspects. In the second part, he devised an interesting method for balancing the two aspects (accuracy and speed) and to estimate natural selection of the best rules for a given environment (viz. ecological rationality of different probabilistic learning rules). Specifically, he considered that we can simulate an intensive care unit (hereafter: ICU) in which doctors are trying to help a patient. There are three options: the doctor either intervenes correctly, wrongly, or - in case she remains uncertain - does not intervene at all. The probability of the patient's survival changes through time and depends on the decision: as time passes, the survival becomes less likely. Similarly, at any point, the correct intervention increases the probability of survival, the wrong intervention decreases it and not intervening at all puts the probability of survival in between the two other options.

Douven demonstrated that using a method of natural optimization can provide another argument in favor of probabilistic inference to the best explanation: although it is a bolder learning rule -- it leads to quicker conclusions and may therefore suffer from inaccuracy -- it is still quick and reliable enough, so that it will typically provide the best trade-off between the two valuable outcomes of reasoning: speed and accuracy. Specifically, in this case he was simulating 200 doctors, 50 learning from diagnostic tests according to each of the 4 learning rules. Then each of them would get 100 simulated patients and would be able to conduct a number of tests on them (100 tests) to diagnose their disease. At the end of a run we can see what the probability of survival was for each of the 200 simulated doctors and the top 100 doctors were duplicated and the bottom 100 erased from the population. This then went on for 100 generations when mostly explanationist doctors remained.

Although his research project included reasoning under uncertainty and an insightful way of balancing the valuable speed and accuracy of reasoning, it did not consider the trustworthiness of information sources and it also did not consider that information (here: diagnostic tests) might be false. Hence, a combination of the research on partial lying (as described above and in [3]) and that of natural optimization for comparing different probabilistic learning

rules (as just described, see also [4]) appeared to be an interesting topic that needed to be tackled.

## 3 METHOD: PART 1

A computer model was created on the basis of other projects described in the previous section. The model included a trust system in which updating was simulated.

Specifically, in the first part we simulated a game of coin bias detection. Agent A is observing the coin and reasoning about its bias, i.e., A is trying to learn how biased the coin is. The simulations consisted of 500 throws of 11 different coins - hypotheses. Each of the 11 coins had its own bias, with probability from 0 to 1 in .1 increments for it to land on heads.

The experiment was repeated a thousand times. In addition to coin throws we also simulated another Agent B, who may be taken as an information source. Agent B was there to provide unreliable and potentially misleading information to A, viz. B is telling A which side the coin is supposedly going to land on, although B does not necessarily provide true information ("B lies to A"). There were also three lists of lies according to the following principles: simple lying (the player states the least probable outcome of the coin, i.e., if the coin is biased to land on heads, agent B will state that it will land on tails), gambler's lying (the player turns the coin secretly and states the opposite of the outcome) and clairvoyant lying (the clairvoyant knows the exact outcome and states the opposite). Bayes's learning rule, Good's learning rule, Popper's learning rule and Explanatory learning rule were used to learn from these data (observations of coins and statements + dynamic trust) to see which belief system update is causing the least epistemic damage. Learning rules offer a way to update the beliefs in the light of the arrival of new relevant evidence. Specifically: Bayes' learning rule requires that the new probability distribution (after learning) corresponds to the prior conditional probability distribution (conditional on the learned piece of evidence and the level of trust in the source). The other three rules are all based on Bayes' but deviate in various ways: the explanatory learning rule adds extra weight to the hypothesis that provides the best explanation. That is, if a coin lands heads 5 times in a row, the best explanation is that it is fully biased towards heads, so this hypothesis gets a probabilistic "push" compared to what Bayes' rule would require. Good's and Popper's rule are similar, except that instead of looking at the best explanation, they award those hypotheses that provide the most confirmatory theories according to measures of confirmation developed by Good and Popper, respectively (see [4] for formal details about the updating rules).

After the simulations are conducted, we then look at the collected data. Specifically, we were interested in the epistemic performance of the rules (how close to the truth they bring an agent) and the speed of convergence towards true hypotheses. To measure how accurate the rules were, we used a measure called Brier's score (or Brier's penalty). The idea is that we can look at mean squared error of the probability distribution: effectively, if a forecast is perfect, the score is 0, and the more off it is, the higher the score, which is also the reason why the score is sometimes called Brier's penalty.

In our simulations, we used it to compare the accuracy of ascribed coin biases. If the simulated coin has a .7 bias to land heads, then ideally our reasoner would assign probability 1 to the hypothesis that the coin is .7 biased. As this is unlikely to happen, we then measure mean squared error of the discrete probability distribution from this ideal outcome. In turn, we can use this to compare the performance of different learning rules.

Similarly, for measuring speed we can simply compare how long it took each reasoner to assign a probability above some threshold value (e.g., above .9) to the true hypothesis about the coin's bias. Note that both the speed of convergence and the accuracy (Brier's score) also depend on the coin that is used in simulations. This is because it is easier to determine the bias of a fully biased coin than of a fair coin: if it always lands on the same side, it is easier to conclude it is fully biased than when it is landing on various sides (note that a fair coin may also land on the same side many times in a row, although such a pattern is more expected from a fully biased coin).

## 4 RESULTS: PART 1

Results mainly show differences in probabilistic learning rules in simple lying when the probability of lying is 1.0 - constant lying, which is also the only part that we are including in this extended abstract. It was found that the best probabilistic learning rule, in this case, is Explanatory learning rule with the lowest Brier penalties (i.e., the lowest inaccuracy). This result has interesting implications: it shows that if the data is misleading, then it may make more sense to use non-Bayesian alternative probabilistic rules.

Note, however, that the accuracy is even greater when we look at control runs, that is, the cases where the information source was ignored, so that the learning agent was merely observing which side the coin landed on without considering what the liar was asserting. This seems to suggest that when we are dealing with unreliable sources of information, it might be best to immediately ignore such sources, e.g., a doctor who notices that her diagnostic tests are unreliable could stop conducting these tests. However, when we look at the speed of convergence, we observe that it makes sense not to ignore such sources if we are also interested in quickly recognizing true hypotheses: control runs were slower than others at least for some of the simulated coins.

Moreover, inference to the best explanation was, contrary to previous research, the fastest but also the most accurate, i.e., it was able to combine both accuracy and speed of reasoning, the two values that previously appeared to be mutually exclusive.

## 5 METHOD: PART 2

The second part of our research followed the approach used by Douven (see [4] and the section on background above). Particularly, we were interested in simulating an ICU with doctors trying to diagnose their patients when the tests are potentially unreliable/misleading. The situation is very similar to part 1: instead of coin biases we deal with diseases that may show some symptom with 0, .1, .2, ..., 1 probability and tests that correspond to partial lying in various lying styles: they are not fully reliable and our doctors estimate the reliability of the tests. We can then look at what the survival rates were for each of the doctor's patients and replicate the top performing half doctors and repeat this for 100 generations. Hence, we combine the research from the first part and the research described in the background section.

## 6 RESULTS: PART 2

The results are interesting: if we look at tests that are constantly misleading (i.e., all of them are unreliable to some degree that needs to be estimated) and if they correspond to what would be akin to simple lying (if it is more likely that a patient has a symptom X than not at the time of the testing, the test will not show the presence of X), then the doctors that infer to the best explanation prevail through generations (see Figure 1 for an example of a simulation and Figure 2 for average percentage of different doctors in our simulations). However, if the tests correspond to being unreliable in what is akin to gambler's or clairvoyant lying we get different results: tests that are unreliable in the gambler's lying style favor both Good's and explanationist reasoning (Figure 3), while those that are like clairvoyant's (always the wrong result) favor Bayes's rule: see Figure 4.

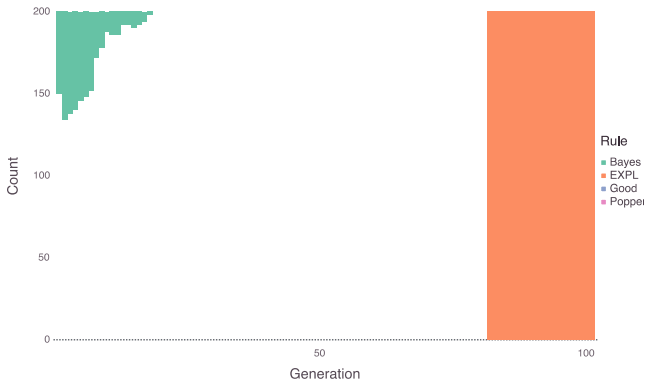



**Figure 1: Example of different agents in a single simulation**

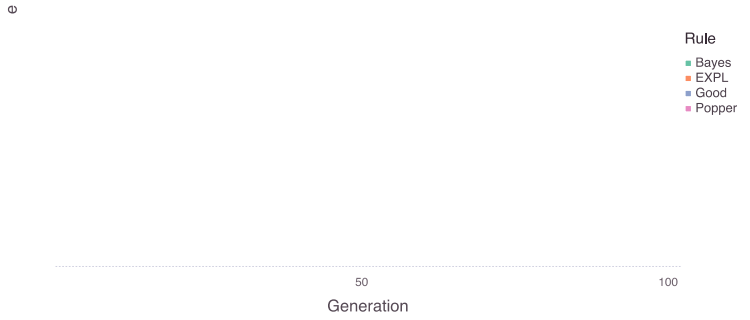



**Figure 2: Average percentage of agents ("simple" lying)**

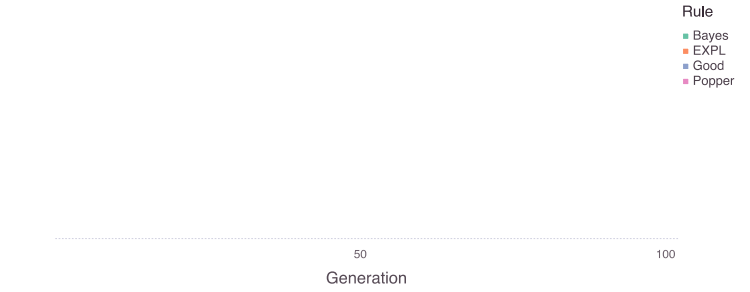



**Figure 3: Average percentage of agents ("gambler's" lying)**

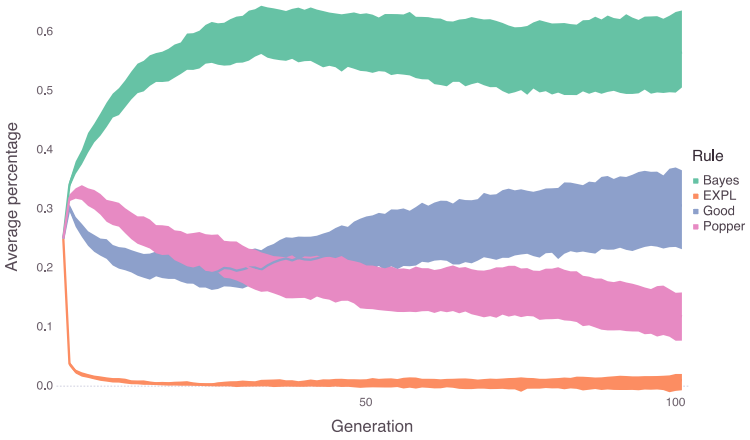



**Figure 4: Average percentage ("clairvoyant" lying)**

## 7 CONCLUSIONS

The results, especially those of Part 2, are very interesting because they suggest that different probabilistic learning rules that have been addressed in literature may be preferable in different situations and preferable in various environments. Ecological rationality then suggests that if we happen to be in an environment with specific features, which we plan to identify in our future research work, then Bayes' rule might be the best way to proceed. Similarly, explanationist learning or Good's or Popper's learning might be preferable in other situations. It remains an open question what features of the information environment determine the choice of a learning rule, but our results suggest that a pluralist approach to learning rules under uncertainty is needed. Our results also provide one possible explanation why we seem to have different reasoning patterns under uncertainty in a descriptive sense, that is, because different environments call for different reasoning strategies. Further research could also provide some insights into pluralist reasoning strategies, i.e., strategy-switching.

## REFERENCES

[1] Igor Douven, 2013. Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. *Philosophical Quarterly*, 63. 428–444. DOI: https://doi.org/10.1111/1467-9213.12032

[2] Borut Trpin and Max Pellert. 2019. Inference to the Best Explanation in Uncertain Evidential Situations. *The British Journal for the Philosophy of Science*, 70, 4. 977-1001. DOI: https://doi.org/10.1093/bjps/axy027

[3] Borut Trpin, Anna Dobrosovestnova, and Sebastian Jakob Götzendorfer. 2020. Lying, more or less: a computer simulation study of graded lies and trust dynamics. *Synthese*, 1-28. DOI: https://doi.org/10.1007/s11229-020-02746-5

[4] Igor Douven. 2020. The ecological rationality of explanatory reasoning. *Studies in History and Philosophy of Science Part A*. 1-14. DOI: https://doi.org/10.1007/s11229-020-02

# Indeks avtorjev / Author index

# Kognitivna znanost

# Cognitive Science

Toma Strle, Borut Trpin, Maša Rebernik, Olga Markič