

Logika, fakty či odmeny

Čo pomáha ľuďom rozpoznať klamstvo?

Richard Priesol, Paulína Jalakšová, Berenika Tužilová,
Matej Lorko, Vladimíra Čavojevová, Jakub Šrol a Zuzana Brokešová

Až tri štvrtiny obyvateľov Slovenska vnímajú šírenie dezinformácií ako závažný problém či priamo bezpečnostnú hrozbu (CEDMO, 2024). Rozhodli sme sa preto preskúmať, aké nástroje pomáhajú ľuďom správne identifikovať dezinformácie a odlíšiť ich od pravdivých informácií. Zamerali sme sa pritom na tvrdenia z geopolitickej oblasti, ktoré sú väčšinou spojené so spoločenskými postojmi, a rovnako aj na tvrdenia z finančnej oblasti, ktoré od nich zvyčajne nezávisia. V dotazníkovom experimente sme porovnali efekt faktických a logických argumentov, a to buď predtým, alebo až potom, čo respondenti prvýkrát ohodnotili príslušné tvrdenia. Ukázalo sa, že všetky testované nástroje zvýšili schopnosť respondentov rozlíšiť pravdivé a nepravdivé tvrdenia o 5 p.b. až 6 p.b. Zároveň sme zistili, že rôzne metódy môžu byť vhodné v rozličných situáciách. Faktické argumenty sú mierne efektívnejšie v prípade priamo vyvrátených dezinformácií, zatiaľ čo logické argumenty sa výrazne ľahšie prenášajú na doposiaľ nevyvrátené dezinformácie. Otestovali sme aj efekt finančných stimulov na úspešnosť v rozlišovaní pravdivých a nepravdivých tvrdení. Odmeny za správne odpovede zvýšili úspešnosť o 2 p.b. až 3 p.b, zatiaľ čo plošné odmeny nemali významný efekt.

Intervencie, ktoré pomáhajú ľuďom identifikovať dezinformácie, sa môžu líšiť obsahovo aj koncepcne. Patria sem jednoduché postrčenia, ako sú napríklad otázky, či užívateľ dôveruje príspevku, ktorý sa chystá zdieľať na sociálnych sieťach, alebo upozornenia na dôležitosť vnímania pravdivosti správ. Ďalej sú to edukačné nástroje, ako napríklad paralelné čítanie a overovanie informácií, mediálna gramotnosť alebo takzvané „očkovanie“, ktoré pomáha pomocou oslabených dezinformácií a manipulačných stratégií vytvárať obranyschopnosť voči takýmto praktikám. Treťou skupinou sú stratégie vyvracania, kam patrí napríklad hodnotenie dôveryhodnosti článkov alebo priame vyvracanie dezinformácií spolu s vysvetlením príslušných súvislostí a faktov (Kozyreva et al., 2024).

V našom experimente sme sa sústredili na priame vyvracanie dezinformácií. Podobne ako Vraga et al. (2020) sme pritom porovnali efekt takejto intervencie pred a po zobrazení dezinformácie, a to pomocou faktických alebo logických argumentov. Tieto stratégie ale môžu pôsobiť rôzne na tvrdenia z rôznych oblastí, keďže ľudia sú všeobecne náchylnejší dôverovať dezinformáciám, ktoré sú v súlade s ich svetonázorom (Ecker et al., 2022). Preto sme sa rozhodli pozrieť sa na tému geopolitiky, ktorá je výrazne spojená so spoločenskými postojmi, keďže priamo súvisí so skupinovú identitou človeka, v porovnaní s omnoho neutrálnejšou témou financií (Tajfel & Turner, 1979). Zároveň môžu byť vyhlásenia ľudí o tom, čomu veria, ovplyvnené aj ich vlastným prospechom (Mercier, 2022). Aj z toho dôvodu sme sa pozreli, či a akým spôsobom zmenia výsledky motivačné odmeny za správne odpovede.

V rámci dotazníkového experimentu sme na vzorke vyše 3 tisíc respondentov porovnali dva spôsoby argumentácie, dve rôzne načasovania a tri druhy odmeňovania za uvedené odpovede. V prvom kole respondenti ohodnotili pravdivosť 12 geopolitických tvrdení, ktoré súvisia so spoločenskými postojmi, a 12 finančných tvrdení, ktoré s nimi nesúvisia. Časť respondentov ešte predtým dostala tri rady, ktoré im mali pomôcť odlíšiť dezinformácie od pravdivých informácií spolu s tromi vyvrátenými dezinformáciami a tromi potvrdenými

pravdivými informáciami. Ďalšia časť dostala rovnakú pomôcku, ale až po ohodnotení všetkých tvrdení. Polovica respondentov pritom dostala logické argumenty, ako napríklad či informácia obsahuje manipulatívny jazyk alebo či z jej šírenia má niekto prospech. Druhej polovici boli naopak predstavené faktické argumenty, napríklad o zdroji danej informácie alebo o postoji príslušných autorít. Kontrolná skupina dostala neutrálny text. V druhom kole, na ktoré boli respondenti pozvaní o dva týždne, sme sa opäť pýtali, ako vnímajú pravdivosť rovnakých tvrdení a tieto výsledky sme navzájom porovnali. Ešte predtým však bola časť respondentov informovaná, že dostanú fixnú sumu, ak zodpovedajú všetky otázky, a druhá časť, že budú odmenení len za správne odpovede. Tretia časť žiadnu informáciu ani odmenu nedostala. Prehľad testovaných nástrojov sa nachádza v Tabuľke 1.

Tabuľka 1: Prehľad použitých intervencií

Argumentácia	Načasovanie	Odmeňovanie
Kontrolná skupina	Kontrolná skupina	Žiadna odmena
Faktické argumenty	Argumenty vopred	Fixná suma
Logické argumenty	Argumenty potom	Za správne odpovede

Box 1: Ďalšie informácie o experimente

Dotazníkový experiment sme uskutočnili prostredníctvom výskumnej agentúry na reprezentatívnej vzorke slovenskej dospeljej populácie. Prvého kola sa zúčastnilo 5 228 respondentov a druhého kola 3 811 z nich, pričom 3 129 dostalo pozvánku do druhého kola presne po dvoch týždňoch od prvého kola a 682 dostalo túto pozvánku skôr a mohli ich tak absolvovať s kratším časovým odstupom. Tieto rozdiely nám umožnili potvrdiť, že účinok intervencií klesá v priebehu času len mierne. V tých skupinách, ktoré dostali pomôcku na rozlíšenie pravdivých informácií a dezinformácií, bola priemerná úspešnosť na úrovni 72 % pri týždňovom odstupe a 68 % pri mesačnom odstupe medzi prvým a druhým kolom.

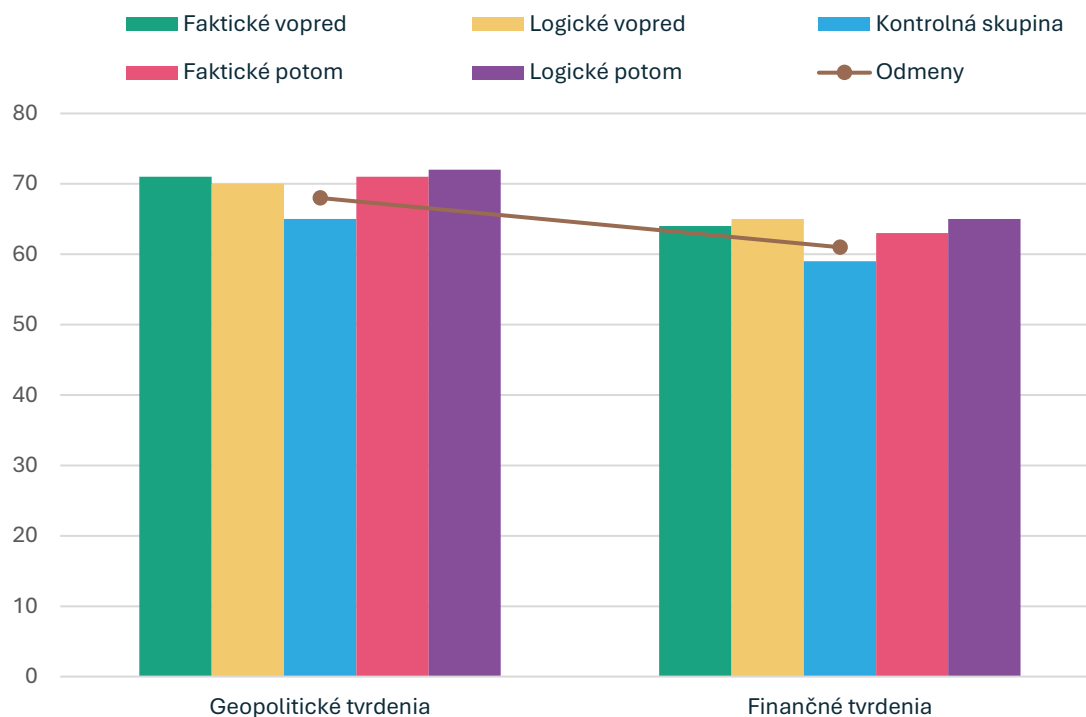
Do experimentu sme vybrali skutočné tvrdenia, ktoré boli v danom čase uverejnené na internete. Pravdivé informácie pritom pochádzali z renomovaných spravodajských médií, zatiaľ čo dezinformácie pochádzali z alternatívnych spravodajských médií a boli vyvrátené organizáciami na overovanie faktov. Respondenti mali tieto tvrdenia ohodnotiť ako pravdivé alebo nepravdivé a následne uviesť, nakoľko sú si istí svojou odpoveďou na 11-bodovej stupnici. Na konci experimentu sme účastníkov informovali o (ne)pravdivosti všetkých tvrdení. Vyplnenie celého dotazníka zabralo približne 17 minút v prvom kole a 21 minút v druhom kole.

V rámci experimentu sme rozdelili respondentov do piatich skupín a každú z nich ešte na tri podskupiny. Jedna skupina nedostala žiadnu pomôcku ako odlišiť dezinformácie od pravdivých informácií a tvorila tak kontrolnú vzorku. Ostatné skupiny dostali tri rady, ktoré ich upozorňovali, na čo si majú pri hodnotení informácií dávať pozor, pričom dve skupiny dostali rady postavené na faktických argumentoch a dve skupiny dostali rady postavené na logických argumentoch. Tieto argumenty boli následne použité na konkrétnych príkladoch, čím sme potvrdili tri pravdivé informácie a naopak vyvrátili tri dezinformácie. Jedna skupina s faktickými argumentami a jedna skupina s logickými argumentami pritom dostali túto pomôcku ešte pred ohodnotením pravdivosti daných tvrdení. Druhá skupina s faktickými

argumentami a druhá skupina s logickými argumentami boli naopak vyzvané, aby najskôr ohodnotili pravdivosť týchto tvrdení a až následne dostali príslušné rady a videli ich využitie. Všetci respondenti hodnotili 12 tvrdení z každej oblasti, z ktorých šesť boli pravdivé informácie, pričom polovica z nich bola potvrdená, a šesť naopak dezinformácie, pričom polovica z nich bola vyvrátená. Tento postup sa opakoval zvlášť v oblasti geopolitiky a zvlášť v oblasti financií, respondenti tak celkovo hodnotili pravdivosť 24 tvrdení. Zároveň sme každú z piatich skupín rozdelili na tri podskupiny podľa udelených odmien. Jedna z nich nedostala žiadnu dodatočnú odmenu, druhá z nich dostala fixnú sumu za zodpovedanie všetkých otázok a tretia z nich bola odmenená len za správne odpovede.

V dotazníku sme sledovali aj rôzne charakteristiky respondentov. V prvom kole sme sa pýtali na ich demografické charakteristiky, ako vek, pohlavie alebo vzdelanie, spoločenské postoje súvisiace so súčasnou geopolitickou situáciou a pocity ohrozenia súvisiace s ekonomickým aj geopolitickým vývojom. V druhom kole sme zisťovali ich dôveru v rôzne inštitúcie, zdroje získavania informácií, konšpiračnú mentalitu, politický cynizmus, politické preferencie, kognitívnu reflexiu a finančnú gramotnosť. Zopakovali sme aj otázky ohľadom spoločenských postojov. Schopnosť rozlíšiť pravdivé informácie a dezinformácie v priemere zlepšovalo vysokoškolské vzdelanie, kognitívna reflexia, finančná gramotnosť, dôvera v medzinárodné inštitúcie a podpora Ukrajiny, a naopak zhoršovala konšpiračná mentalita, dôvera v politické inštitúcie a podpora Ruska.

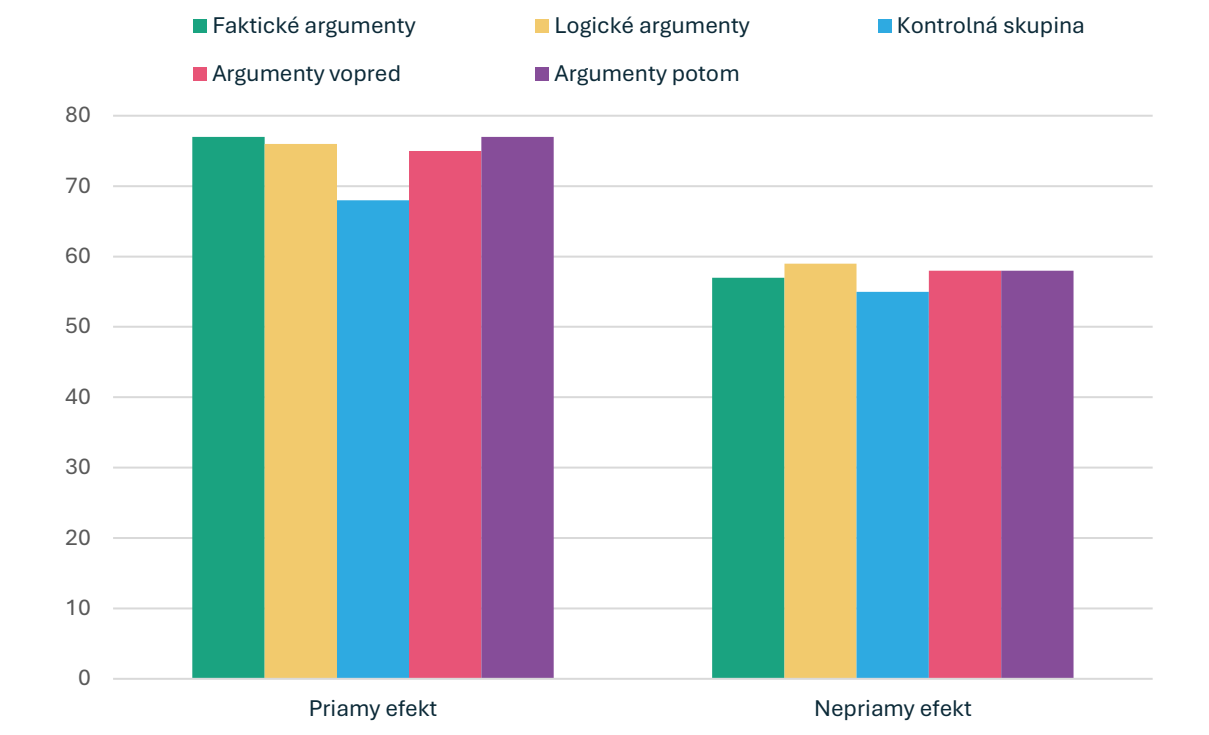
Graf 1: Správne identifikované tvrdenia podľa skúmaných oblastí (%)



Faktické aj logické argumenty, či už boli poskytnuté pred alebo po prvom ohodnotení vybraných tvrdení, zvýšili mieru ich správnej identifikácie o 5 p.b. až 6 p.b. Efekt použitých argumentov bol teda významný, avšak podobný medzi jednotlivými skupinami. Odmeny za správne odpovede zvýšili úspešnosť respondentov o 2 p.b. až 3 p.b., a to napriek tomu, že samotnému textu či odpovedaniu nevenovali viac času. Plošné odmeny za zodpovedanie

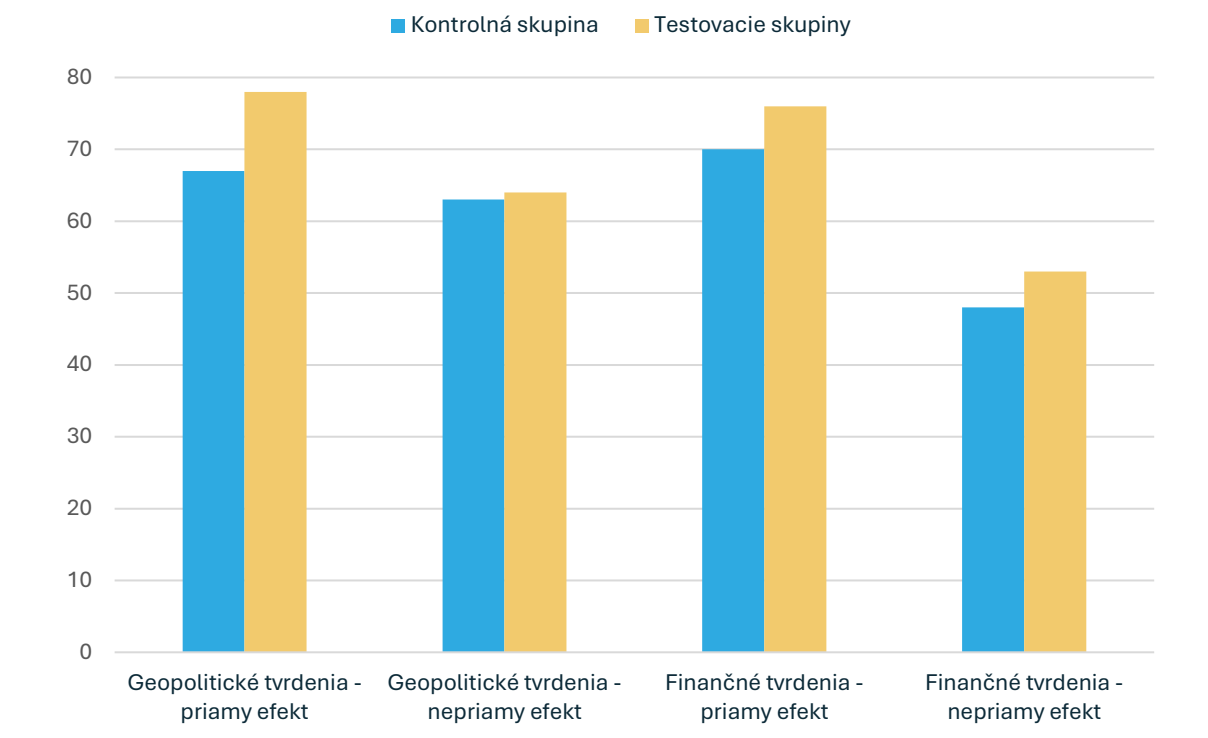
všetkých otázok pritom nemali na úspešnosť respondentov významný efekt. Respondenti boli okrem toho úspešnejší v identifikovaní pravdivosti geopolitických tvrdení v porovnaní s tými finančnými. To môže mať viacero vysvetlení, ktoré sú čiastočne založené na predpoklade, že s témou geopolitiky sa ľudia stretávajú častejšie než s témou financií a poznajú tak viac podobných tvrdení, o ktoré sa môžu pri rozhodovaní oprieť. Vďaka tomu si podľa teórie schém (Bartlett, 1932) vytvárajú viac schém pre geopolitické ako finančné koncepty. Geopolitické informácie sa navyše častejšie podávajú formou príbehu, ktorý sa ľahšie spracováva, zatiaľ čo finančné informácie zvyčajne vyžadujú určitú dávku abstraktného myslenia, čo zvyšuje ich kognitívnu náročnosť. Tento rozdiel môže na základe modelu pravdepodobnosti elaborácie (Petty & Cacioppo, 1986) viesť k povrchnejšiemu spracovaniu finančných tém. Napokon, podľa kultúrnej kognitívnej teórie (Kahan & Braman, 2006), venujú ľudia viac pozornosti geopolitickým témam aj preto, že súvisia s ich kultúrnymi hodnotami. Mieru správne identifikovaných tvrdení v závislosti od skúmaných oblastí vidíme v Grafe 1.

Graf 2: Správne identifikované tvrdenia podľa skúmaných efektov (%)



Fakty boli mierne účinnejšie pri priamom vyvracaní, logika naopak pri tvrdeniach, ktoré neboli priamo vyvrátené. Faktické argumenty zvýšili mieru správnej identifikácie najmä tých tvrdení, ktoré boli priamo potvrdené či vyvrátené, a to o 9 p.b. Logické argumenty dosiahli mierne slabší efekt na úrovni 8 p.b. Takáto argumentácia však účinkovala oveľa lepšie na nové tvrdenia, ktorých pravdivosť respondenti nemali potvrdenú či vyvrátenú. V tomto prípade logické argumenty zvýšili mieru ich správnej identifikácie o 4 p.b, zatiaľ čo faktické argumenty len o 2 p.b. Pre tie informácie, ktoré boli priamo potvrdené alebo vyvrátené, bola intervencia mierne účinnejšia, ak bola načasovaná až potom, čo mali respondenti možnosť dané tvrdenia sami ohodnotiť. Následné vyvrátenie zvýšilo mieru ich správnej identifikácie o 9 p.b, zatiaľ čo preventívne vyvrátenie iba o 7 p.b. V prípade nových tvrdení nemalo načasovanie intervencie žiadny efekt na jej účinnosť. Mieru správne identifikovaných tvrdení v závislosti od skúmaných efektov vidíme v Grafe 2.

Graf 3: Správne identifikované tvrdenia podľa oblastí a efektov (%)



V prípade priameho vyvracania mali argumenty väčší efekt na geopolitické ako finančné tvrdenia, v prípade nových tvrdení to bolo presne naopak. Poskytnutie argumentov zvýšilo mieru správnej identifikácie tvrdení, ktoré boli priamo potvrdené alebo vyvrátené, a to o 11 p.b. v oblasti geopolitiky a 6 p.b. v oblasti financií. Je teda možné, že priame vyvracanie rezonovalo viac pri tvrdeniach spojených so skupinovú identitou. Naopak, pri použití rovnakej pomôcky na nové tvrdenia, ktoré neboli priamo potvrdené alebo vyvrátené, nemali tieto argumenty žiadny efekt v oblasti geopolitiky, pričom v oblasti financií zvýšili mieru ich správnej identifikácie až o 5 p.b. To môže naznačovať, že zatiaľ čo v oblastiach súvisiacich so skupinovú identitou je veľmi náročné dosiahnuť prenos nadobudnutých poznatkov na nové informácie, v oblastiach nesúvisiacich so skupinovú identitou je takýto prenos výrazne jednoduchší. Mieru správne identifikovaných tvrdení v závislosti od oblastí a efektov vidíme v Grafe 3.

V našom experimente sme zistili, že vyvracanie a potvrdzovanie informácií má významný efekt na ich správnu identifikáciu, pričom rôzne metódy môžu byť vhodné v rozličných situáciách. Ak je našim hlavným cieľom potvrdiť alebo vyvrátiť konkrétne tvrdenia, ako mierne účinnejšia metóda sa javí poskytnutie dodatočných faktov o zdroji týchto informácií a postoj príslušných autorít, a to ideálne ihneď po kontakte s nimi. Naopak, ak je našim hlavným cieľom potvrdzovať alebo vyvracať dosiaľ nešpecifikované tvrdenia, lepšie výsledky prinášajú logické argumenty nabádajúce k zamysleniu, či informácia obsahuje manipulatívny jazyk alebo či z jej šírenia má niekto prospech. Finančná odmena takisto zvýšila pomer správne identifikovaných tvrdení, a to pravdepodobne prostredníctvom zvýšenia motivácie respondentov. To naznačuje, že aj účinné upriamenie pozornosti na potrebu správnej identifikácie informácií môže priniesť významné zlepšenie v rozoznávaní ich pravdivosti. Napokon, prenos nadobudnutých poznatkov na nové informácie je výrazne jednoduchší v prípade takých oblastí, ktoré nie sú spojené so skupinovú identitou človeka.

Podakovanie

Tento výskum bolo možné uskutočniť vďaka finančnej podpore Nadácie NBS z výskumného projektu GV-2023-7-P12.

Literatúra

Bartlett, F. C. (1932): Remembering: A Study in Experimental and Social Psychology. Cambridge, UK: Cambridge University Press.

Central European Digital Media Observatory (2024): Three-Quarters of Slovaks Perceive the Spread of Disinformation as a Threat to Slovakia's Security. <https://cedmohub.eu/three-quarters-of-slovaks-perceive-the-spread-of-disinformation-as-a-threat-to-slovakias-security>

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K. & Amazeen, M. A. (2022): The Psychological Drivers of Misinformation Belief and Its Resistance to Correction. *Nature Reviews Psychology*, Vol. 1, No. 1.

Kahan, D. M. & Braman, D. (2006): Cultural Cognition and Public Policy. *Yale Law & Policy Review*, Vol. 24, No. 1.

Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Lewandowsky, S. & Hertwig, R. (2024): Toolbox of Individual-Level Interventions against Online Misinformation. *Nature Human Behaviour*, Vol. 8, No. 8.

Mercier, H. (2022): Not Born Yesterday: The Science of Who We Trust and What We Believe. Princeton, NJ: Princeton University Press.

Petty, R. E. & Cacioppo, J. T. (1986): *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York, NY: Springer-Verlag.

Tajfel, H. & Turner, J. C. (1979): An Integrative Theory of Intergroup Conflict. In: *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks/Cole.

Vraga, E. K., Kim, S. C., Cook, J. & Bode, L. (2020): Testing the Effectiveness of Correction Placement and Type on Instagram. *The International Journal of Press/Politics*, Vol. 25, No. 4.