



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Aceitação e Correção de Desinformação: O Papel da Credibilidade da Fonte

Carolina Cabral Pereira Mendes

Mestrado em Psicologia Social e das Organizações

Orientadora: Margarida Vaz Garrido
Professora Associada com Agregação
Iscte – Instituto Universitário de Lisboa

Coorientadora: Magda Saraiva
Investigadora Auxiliar
William James Center for Research, Ispa – Instituto
Universitário

Julho, 2025



CIÊNCIAS SOCIAIS
E HUMANAS

Departamento de Psicologia Social e das Organizações

Aceitação e Correção de Desinformação: O Papel da Credibilidade da Fonte

Carolina Cabral Pereira Mendes

Mestrado em Psicologia Social e das Organizações

Orientadora: Margarida Vaz Garrido
Professora Associada com Agregação
Iscte – Instituto Universitário de Lisboa

Coorientadora: Magda Saraiva
Investigadora Auxiliar
William James Center for Research, Ispa – Instituto Universitário

Julho, 2025

Agradecimentos

A realização desta tese representa a concretização de uma importante etapa na minha vida académica e pessoal, e não teria sido possível sem o apoio, incentivo e colaboração de várias pessoas a quem quero expressar o meu profundo agradecimento.

Aos meus pais, pelo carinho e apoio constante em todos os momentos. E à Bia, a minha melhor amiga e alma gémea, por seres dos meus maiores exemplos de perseverança e disciplina.

À Marta, que está lá desde o primeiro dia e que ouve todos os meus desabafos e preocupações. Sou uma melhor pessoa por te ter na minha vida.

À Mafalda, Inês e Joana A., que iniciaram este caminho comigo, e a quem agora chamo amigas, agradeço as vossas partilhas e companheirismo. Foi muito bom fazer este percurso convosco.

À Joana D., que acompanhou o início da minha jornada profissional, e que sempre me incentivou a não desistir. Tens sido um exemplo e uma grande conselheira.

À Cláudia, que fez mais pela minha saúde mental do que alguma vez serei capaz de colocar por escrito. Lembro-me frequentemente dos teus ensinamentos e levo-os para a vida.

Aos restantes familiares e amigos, agradeço pelo interesse que sempre demonstraram neste (demorado) trabalho e por nunca me deixarem sentir sozinha neste que acaba por ser um projeto tão solitário.

À Sara, deixo também uma palavra de gratidão pela disponibilidade e paciência com que me auxiliou e esclareceu tantas questões. Este trabalho não teria sido possível sem a sua preciosa ajuda e sentido crítico.

Termino agradecendo às minhas orientadoras, Professora Margarida Garrido e Professora Magda Saraiva pelo tempo que dedicaram a este projeto, pela paciência, e acima de tudo pelos valiosos comentários e observações. A exigência e assertividade de ambas foi essencial para a concretização deste trabalho.

Resumo

As redes sociais têm permitido a rápida disseminação de conteúdos não verificados, o que facilita a propagação de informações falsas sobre diferentes temas, nomeadamente, na área da saúde. Esta tendência representa um risco significativo para a saúde pública, pois pode levar à adoção de comportamentos de risco (Bridgam et al., 2020; Islam et al., 2020).

A presente dissertação procura examinar o papel da credibilidade da fonte na aceitação e correção de desinformação online, no contexto da saúde. Para tal, realizámos um estudo experimental, com o intuito de explorar de que forma a perceção da credibilidade da fonte influencia os efeitos de ilusão de verdade e de ilusão de falsidade.

Os resultados revelaram que a credibilidade da fonte teve um impacto significativo nos julgamentos de verdade, na medida em que as frases partilhadas por fontes consideradas credíveis foram julgadas como mais verdadeiras. No que se refere à interação entre a credibilidade da fonte e a avaliação das frases repetidas na fase de re-exposição observou-se um padrão semelhante. No entanto, apenas a diferença entre frases partilhadas por fontes credíveis e sem fonte foi significativa. O efeito de ilusão de verdade não foi consistentemente observado, e o efeito de ilusão de falsidade emergiu na direção oposta à esperada.

Não obstante as limitações identificadas, os resultados reforçam a necessidade de se considerar a origem da informação no combate à desinformação, sobretudo em áreas críticas como a saúde pública.

Palavras-chave

Ilusão de verdade; Ilusão de falsidade; Credibilidade da fonte; Correção de desinformação

Códigos e Classificações da APA

2340 Processos Cognitivos

3040 Perceção Social e Cognição

Abstract

Social media has enabled the rapid dissemination of unverified content, facilitating the spread of false information on various topics, particularly in the health sector. This trend poses a significant threat to public health, as it can lead to the adoption of risk behaviors (Bridgam et al., 2020; Islam et al., 2020).

This dissertation aims to examine the role of source credibility in the acceptance and correction of online health misinformation. To this end, we conducted an experimental study to explore how the perception of source credibility influences the illusion of truth and the illusion of falseness.

The results revealed that source credibility had a significant impact on participants' truth judgments, with statements shared by credible sources being considered more truthful. A similar pattern was observed regarding the interaction between source credibility and the evaluation of repeated statements during the re-exposure phase. However, only the difference between statements shared by credible sources and those without a source was significant. The illusion of truth effect was not consistently observed, and the illusion of falseness emerged in the opposite direction to that expected.

Despite the identified limitations, the results underscore the importance of considering the origin of information in the fight against misinformation, particularly in critical areas such as public health.

Keywords:

Illusion of truth; Illusion of falseness; Source credibility; Misinformation correction

APA Classification Categories and Codes

2340 Cognitive Processes

3040 Social Perception & Cognition

Índice

Introdução.....	1
CAPÍTULO 1. Enquadramento Teórico	3
1.1. Desinformação na área da saúde	3
1.2. Fatores que favorecem a aceitação da desinformação	4
1.3. Paradigma da desinformação	8
1.4. Paradigma do efeito da ilusão de verdade (e ilusão de falsidade)	10
1.5. Estratégias de correção da desinformação	11
1.6. Objetivos e hipóteses	12
CAPÍTULO 2. Estudo Piloto	15
2.1. Método.....	15
2.2. Resultados.....	16
CAPÍTULO 3. Estudo Experimental.....	19
3.1. Método.....	19
3.2. Resultados.....	23
CAPÍTULO 4. Discussão	27
4.1. Principais resultados	27
4.2. Limitações e sugestões para estudos futuros	29
Conclusão	30
Referências Bibliográficas.....	31
Anexos.....	41

Introdução

Os termos desinformação e *fake news* ganharam particular destaque aquando das eleições presidenciais americanas, em 2016. Durante esse período, a rápida circulação de informações falsas nas redes sociais atraiu a atenção de jornalistas e investigadores, e foram vários os autores que se debruçaram sobre o tema (e.g., Allcott & Gentzkow, 2017; Grinberg et al., 2017).

Porém, a desinformação prolifera em vários domínios que não só a política. De facto, dada a velocidade a que a informação se propaga atualmente nas redes sociais, este fenómeno é transversal a quase todos os temas, nomeadamente a saúde.

Por desinformação entende-se informação falsa ou incorreta, que é divulgada sem que haja, necessariamente, a intenção de enganar (Lewandowsky et al., 2017), ao contrário das *fake news* que envolvem a partilha deliberada de informação falsa com o intuito de enganar os leitores e manipular a opinião pública (Lazer et al., 2018).

No contexto das redes sociais, a desinformação resulta num aumento das interações por parte dos utilizadores e, consequentemente, a maiores receitas geradas (Allcott & Gentzkow, 2017; Papadogiannakis et al., 2023). Estudos anteriores têm observado que estas informações falsas se propagam nas redes sociais mais rapidamente do que as verdadeiras, embora este fenómeno ainda não tenha sido estudado de forma aprofundada em todas as plataformas – apenas no X (antigo Twitter) e Facebook (e.g., Guess et al., 2019; Moravec et al., 2018).

A desinformação online acarreta, inevitavelmente, consequências individuais e sociais (Lewandowsky et al., 2012), que se fazem sentir sobretudo ao nível dos padrões de consumo dos indivíduos e nas áreas da política, finanças e saúde (Bovet & Makse, 2019; Greifeneder et al., 2020).

Atualmente, assiste-se a uma preocupação crescente com as consequências da desinformação na área da saúde, devido à facilidade de propagação de conteúdos incorretos muitas vezes “mascarados” de evidências científicas (Budak et al., 2024; Neyazi et al., 2021) e cuja veracidade, muitos indivíduos não têm capacidade de julgar adequadamente.

Um exemplo recente de desinformação em massa na área da saúde ocorreu aquando da pandemia da COVID-19, durante a qual circularam inúmeras teorias da conspiração e informações incorretas acerca da origem do vírus e da eficácia de determinados medicamentos no combate ao mesmo (Rocha et al., 2021). Isto resultou numa maior resistência à adoção de medidas preventivas, como o uso de máscaras e a adesão à vacinação (Lee et al., 2022; Zimmerman et al., 2023), assim como o aumento de situações de violência e preconceito contra a população chinesa (Koller et al., 2021).

Ainda assim, já no passado eram amplamente divulgadas informações incorretas sobre a saúde e bem-estar e, como se tem observado, não é necessário um evento em grande escala (como a pandemia COVID-19), para que a desinformação se propague. Aliás, existem informações que começaram a ser divulgadas há dezenas de anos atrás e que, ainda hoje estão presentes e são temas recorrentes em fóruns de discussão online. Exemplo disso é a (falsa) ideia de que a vacinação contra o sarampo na infância está associada a perturbações do espectro do autismo (Wakefield et al., 1998). Apesar de esta ideia não ter sido inicialmente difundida na internet, com o passar dos anos e o crescimento das redes sociais, começou a circular em diferentes fóruns e, ainda hoje, não obstante os vários desmentidos, permanece como uma falsa informação em constante circulação (Boom et al., 2018).

Em resumo, a circulação de desinformação nas redes sociais tem a capacidade de influenciar tópicos importantes, como eleições, atitudes dos cidadãos face a políticas públicas e decisões da população sobre cuidados de saúde (Calvillo et al., 2021). Deste modo, o impacto das redes sociais na aceitação da desinformação revela-se um tema preocupante no que toca à saúde pública, na medida em que pode reduzir a eficácia de programas, campanhas e iniciativas que visam a saúde e bem-estar dos cidadãos (Southwell et al., 2019). A presente dissertação tem como objetivo aprofundar esta temática e especificamente investigar o papel da credibilidade da fonte da informação na aceitação e correção de desinformação no contexto da saúde.

O Capítulo 1 explora o conceito de desinformação, com base nos fatores que, à data, melhor explicam a aceitação de desinformação e as estratégias que se têm revelado mais eficazes na sua correção. Será também feito um enquadramento teórico dos paradigmas utilizados neste estudo.

No Capítulo 2, é apresentado o estudo piloto que permitiu selecionar os materiais mais adequados ao estudo principal e no Capítulo 3 apresenta-se um estudo experimental que testa o papel da credibilidade da fonte na aceitação e correção de desinformação.

No Capítulo 4, são discutidos os principais resultados, assim como as limitações deste trabalho e algumas sugestões para estudos futuros.

CAPÍTULO 1.

Enquadramento Teórico

1.1. Desinformação na área da saúde

A desinformação corresponde a informação falsa ou incorreta, divulgada sem que exista, necessariamente, o intuito de enganar (Lewandowsky et al., 2012) e que pode levar à distorção de informação apresentada previamente (Scheufele & Krause, 2019). A desinformação pode resultar da má interpretação de factos e desempenha um papel importante na distorção das percepções dos indivíduos, particularmente no contexto político, social e da saúde (Bode & Vraga, 2018). Quem divulga este tipo de informações, nem sempre tem a percepção de que se trata de informações incorretas e acredita até que pode estar a ajudar outros indivíduos partilhando conhecimento relevante (Vraga & Bode, 2017).

As fontes mais comuns de desinformação na área da saúde são a má interpretação de estudos científicos (West & Bergstrom, 2021), interesses comerciais (Lewandowsky et. al., 2017), informação médica desatualizada (Bode & Vraga, 2018), estudos científicos pouco consistentes e/ou fraudulentos (Ioannidis, 2005), e alegações pseudocientíficas (Blaskiewicz, 2013).

No contexto das redes sociais, assiste-se a uma preferência por títulos chamativos e conteúdos resumidos que possam ser consumidos rapidamente e sem grande esforço, o que leva frequentemente a um encurtamento das notícias e a uma certa distorção dos factos (Wu et al., 2020). Para além disso, importa destacar que conteúdos que despoletem fortes respostas emocionais (como o medo ou a raiva) são mais prováveis de se tornarem virais (Vosoughi et al., 2018). A área da saúde constitui um campo propício a debates acesos e partilhas de opiniões, suportados por plataformas digitais e entidades que lucram com esse tipo de conteúdos e consequentes reações (Scherer et al., 2020).

São inúmeras as consequências da desinformação na área da saúde e o artigo de Wakefield (1998) é um bom exemplo de como até estudos alegadamente científicos podem dar origem a mitos que se propagam ao longo de gerações. Este artigo sugere que a vacinação contra o sarampo na infância poderia causar autismo, o que levou muitos pais a decidir não vacinar os seus filhos e ao consequente reaparecimento de doenças que até essa altura estavam altamente controladas (Godlee et al., 2011). Ainda que as informações contidas no estudo de Wakefield tenham sido repetidamente corrigidas, muitos indivíduos ainda acreditam na ligação entre a vacinação na infância e as perturbações do espectro do autismo, ilustrando as consequências e a longevidade da desinformação.

1.2. Fatores que favorecem a aceitação da desinformação

A aceitação de desinformação não tem uma causa única e a literatura tem identificado recorrentemente alguns fatores que parecem estar relacionados com este fenómeno, como ilustramos de seguida.

Falta de literacia em saúde

A literacia em saúde é definida como a capacidade de obter, processar e compreender informações básicas sobre saúde necessárias a uma tomada de decisão apropriada e informada (Nutbeam, 2000). Assim, a literacia em saúde afeta a forma como os indivíduos interpretam informações médicas, seguem planos de tratamento, e utilizam os sistemas de cuidados de saúde (Berkman et al., 2011).

A falta de conhecimentos nesta área também faz com que determinadas populações não adotem medidas de saúde preventivas com a frequência necessária (Berkman et al., 2011) e não sejam capazes de gerir algumas condições de saúde de forma adequada (Bailey et al., 2014), levando a um maior número de hospitalizações e consequente sobrecarga dos sistemas de cuidados de saúde.

Por exemplo, Mitchell et al. (2012), mostraram que pacientes com baixos níveis de literacia em saúde, e que tinham sido submetidos a cirurgias cardiovasculares, tinham uma maior probabilidade de serem readmitidos no hospital nos trinta dias a seguir à operação. Ao explorar as causas deste fenómeno, os autores perceberam que estes pacientes revelavam uma maior dificuldade em compreender as indicações dadas no momento da alta hospitalar, pelo que não seguiam os cuidados pós-operatórios adequados.

Este tipo de cenários é amplificado pela facilidade e rapidez no acesso a informações falsas através das redes sociais, levando muitas pessoas a procurar (e adotar) tratamentos alternativos e “curas” sem qualquer evidência científica, assim como a abandonar planos de tratamento prescritos por profissionais de saúde (Betsch et al., 2010; Davis et al., 2006).

Viés da confirmação

Aliada à baixa literacia em saúde, assiste-se também a uma tendência das pessoas para procurarem informações que vão ao encontro das suas crenças, ignorando ou rejeitando informações que as contradizem – viés da confirmação (Nickerson, 1998). Assim, a informação (verdadeira ou falsa) que é consistente com o conhecimento prévio tende a ter uma maior aceitação (Judd & Kulik, 1980; Petty et al., 1997) e torna-se altamente resistente à mudança (Lewandowsky et al., 2012).

Por detrás deste enviesamento está um fenómeno explorado por Festinger (1957) – dissonância cognitiva - que se prende com o desconforto sentido pelos indivíduos quando se

deparam com crenças contraditórias. Assim, tendem a preferir informações que suportam as suas crenças atuais, independentemente da veracidade das mesmas.

No contexto das redes sociais, em que existem algoritmos que personalizam os conteúdos apresentados aos utilizadores com base nas suas interações (Gillespie, 2014), os indivíduos acabam por ser maioritariamente expostos a informações concordantes com as suas perspetivas e ao interagirem positivamente com as mesmas (e.g., através de *likes*, comentários ou histórico de pesquisa), continuam a ser-lhes apresentadas informações e conteúdos idênticos, o que pode reforçar crenças erradas. Por outras palavras, a natureza interativa das redes sociais também alimenta este ciclo de *feedback*; quando as pessoas partilham ou seguem um perfil alinhado com o seu ponto de vista, estão a sinalizar ao algoritmo que aquele tipo de conteúdos lhes interessa. Consequentemente, ser-lhes-ão apresentados mais conteúdos da mesma natureza, o que pode contribuir para a divulgação e enraizamento de desinformação (Del Vicario et al., 2016b; Vosoughi et al., 2018).

Além disso, mesmo que posteriormente confrontados com informações factuais corretas, os indivíduos tendem a revelar alguma resistência à mudança de opinião (Nyhan & Reifler, 2010), sobretudo em ambientes onde são repetidamente expostos a desinformação.

Traços de personalidade

Alguns traços de personalidade também têm sido associados à propensão para acreditar em informações falsas e à dificuldade de julgar adequadamente a informação (Ahmed & Tan, 2022; Piksa et al., 2022), responder à sua correção e avaliar a credibilidade da sua fonte. O NEO PI – R – Inventário de Personalidade (Costa & McCrae, 1992) tem sido o instrumento mais utilizado neste tipo de estudos, no que toca à avaliação das cinco principais dimensões da personalidade humana (neuroticismo, extroversão, abertura à experiência, amabilidade e conscienciosidade).

A abertura à experiência, por exemplo, caracteriza-se pela curiosidade intelectual e pela imaginação (McCrae & John, 1992) e está associada a uma maior flexibilidade cognitiva e à vontade de adquirir novas informações. A investigação tem mostrado que os indivíduos com uma maior abertura à experiência têm maior probabilidade de apresentarem um pensamento crítico e questionarem as fontes da informação e, em resultado, menor probabilidade de acreditar em desinformação (DeYoung et al., 2007; Pennycook et al., 2020).

Por outro lado, a conscienciosidade reflete-se numa maneira de ser organizada, responsável e cuidadosa (John & Srivastava, 1999) – que pode ser protetora no que respeita à aceitação de desinformação - mas as evidências encontradas na literatura são contraditórias (Lin et al., 2023). Por um lado, indivíduos altamente conscienciosos tendem a partilhar menos informações falsas, devido à atenção ao detalhe que demonstram e à preferência que dão a informações

factualmente corretas (Calvillo et al., 2024). Por outro lado, estes indivíduos também podem revelar uma maior tendência para obedecer a figuras de autoridade e seguir regras e crenças tradicionais (Jost et al., 2003; Sibley & Duckitt, 2008), que podem não estar corretas.

De igual modo, a amabilidade também tem dado origem a resultados por vezes contraditórios. Indivíduos com altos níveis de amabilidade são descritos como calorosos, cooperativos e com tendência para evitar conflitos (Graziano & Tobin, 2002), o que faria supor que apresentassem também uma predisposição para considerar vários pontos de vista (McCrae & Costa, 1997). Contudo, alguns estudos mostram que estes indivíduos tendem a partilhar desinformação em alguns contextos particulares, sobretudo quando a informação tem uma grande carga emocional ou apela à empatia (Flynn et al., 2017; Van Prooijen & Douglas, 2018).

Também a extroversão, associada à sociabilidade, assertividade, entusiasmo e desejo de envolvimento social (Möttus & Rozgonjuk, 2021), pode ter um papel relevante na aceitação de desinformação. Indivíduos mais extrovertidos tendem a relacionar-se mais com outras pessoas (*online* e presencialmente), o que os torna mais expostos a vários tipos de informação, nomeadamente desinformação (Pennycook et., 2020). Para além disso, indivíduos mais extrovertidos tendem a valorizar a harmonia social e evitam discórdias, o que também os torna mais propícios a aceitar informações alinhadas com o ponto de vista dos seus pares, sem qualquer verificação ou análise dos factos (Guess et al., 2019; Osmundsen et al., 2021).

Por fim, o neuroticismo tem sido o traço de personalidade mais relacionado com a aceitação de desinformação, por estar associado a elevados níveis de instabilidade, ansiedade e respostas emocionais negativas ao stress (Costa & McCrae, 1992; Lahey, 2009). Indivíduos com mais traços neuróticos parecem ser mais vulneráveis a acreditar em desinformação, devido à sua tendência para reagir mais intensamente a conteúdos com uma maior carga emocional (Calvillo et al., 2024). Assim, informações que induzam medo, raiva ou ansiedade – como teorias da conspiração ou mitos relacionados com saúde (Swami et al., 2014) – despertam reações mais intensas nestes indivíduos, o que dificulta a sua análise crítica da informação.

Credibilidade e confiança na fonte

A credibilidade da fonte é descrita através da confiança percebida e o nível de *expertise* da fonte da informação (Metzger & Flanagin, 2013). A confiança percebida diz respeito à honestidade e integridade que a fonte transmite ao leitor, enquanto o nível de *expertise* corresponde aos conhecimentos e competências da fonte numa determinada área (Hovland et al., 1953).

Algumas fontes de informação, como instituições de ensino ou agências governamentais, são vistas como mais credíveis por terem, regra geral, normas e padrões rigorosos que visam a

imparcialidade, precisão e confiabilidade dos dados que partilham (Kovach & Rosentiel, 2014). De facto, as instituições de ensino (e.g., universidades), centros de investigação e profissionais associados aos mesmos, são vistos como fontes de informação fidedignas, uma vez que os dados que partilham estão sujeitos a uma revisão dos seus pares e códigos de conduta (Kovach & Rosentiel, 2014), que visam a objetividade e transparência (Singer, 2015).

Ainda assim, apesar de estas entidades continuarem a ser fontes de referência, o aparecimento dos chamados criadores de conteúdos (ou *influencers*) nas redes sociais veio alterar o conceito de autoridade e confiabilidade das fontes de informação (Metzger & Flanagin, 2013). Paralelamente, fez também com que os utilizadores comessem a fazer julgamentos da informação online com base em pistas mais superficiais, como a popularidade ou a aparência da fonte e das suas publicações (De Veirman et al., 2017; Djafarova & Rushworth, 2017).

Além disso, familiares e amigos também têm a capacidade de influenciar a nossa aceitação de desinformação, pois dada a sua familiaridade e relação próxima, são automaticamente percebidos como fontes credíveis. Efetivamente, as pessoas têm mais probabilidade de acreditar e partilhar informações que advenham de alguém que conhecem pessoalmente do que de outras fontes (Del Vicario et al., 2016a; Metzger et al., 2010), mesmo sem evidências para tal. Para além disso, duvidar da credibilidade de uma informação partilhada por um ente querido pode ser visto como um desafio à sua capacidade de julgamento ou confiança e, por consequência, gerar conflitos (Bessi et al., 2015). Assim, tendemos a ignorar incoerências factuais ou a corrigir familiares ou amigos, perpetuando a divulgação de desinformação (Bessi et al., 2015; Pennycook & Rand, 2018).

Carga emocional da informação

As emoções também influenciam a forma como os indivíduos reagem, processam e partilham informações (Brady et al., 2017), e aquelas que despoletam fortes reações emocionais têm maior probabilidade de serem recordadas e partilhadas (Lang et al., 1997).

O medo, a raiva, a surpresa e a alegria costumam ser as emoções mais relevantes no contexto da desinformação (Pennycook et., 2020) por ativarem áreas cerebrais associadas à atenção e memória (Kensinger & Schacter, 2008) e por tornarem os indivíduos mais suscetíveis a aceitar a informação sem verificar a sua veracidade (Forgas, 1995). Isto acontece pelo facto de algumas emoções limitarem a capacidade cognitiva de pensar criticamente e propiciarem a utilização de atalhos mentais (e.g., heurísticas) aquando da avaliação da informação (Forgas, 1995).

As redes sociais amplificam os efeitos da carga emocional de algumas informações na medida em que certos conteúdos, ao despoletarem sentimentos de medo, raiva, nojo ou

entusiasmo, tendem a tornar-se virais mais rapidamente do que informações baseadas em factos ou neutras (Berger & Milkman, 2012). Isto deve-se também ao facto de as redes sociais (através das suas funcionalidades – e.g., partilha, comentários) despoletaram um rápido contágio emocional, comparativamente com aquilo que sucede quando as informações são partilhadas boca a boca (Berger & Milkman, 2012; Kramer et al., 2014).

Para além disso, como referido anteriormente, os algoritmos das redes sociais são desenhados de forma a priorizarem os conteúdos que geram *engagement* e interações por parte dos utilizadores. Habitualmente, estes conteúdos virais são precisamente os que despoletam mais reações e emoções mais fortes nos indivíduos (Bakir & McStay, 2018; Lazer et al., 2018).

Os resultados da breve revisão de literatura apresentada revelam que a aceitação da desinformação está profundamente associada a diversos fatores cognitivos, afetivos e sociais (Lewandowsky et al., 2012). No presente estudo vamos debruçar-nos sobre a credibilidade da fonte.

São também vários os paradigmas utilizados para estudar este fenómeno – nomeadamente o paradigma da desinformação (Loftus & Palmer, 1974) e o paradigma do efeito da ilusão de verdade (Bacon, 1979; Hasher et. al., 1977) que expomos brevemente na secção seguinte.

1.3. Paradigma da desinformação

O paradigma da desinformação foi desenvolvido nos anos 70, por Elizabeth Loftus e colaboradores e corresponde à distorção da memória para informações ou eventos anteriormente codificados, através da introdução deliberada de informação errada/enganosa (Loftus & Palmer, 1974). Este paradigma tem três fases distintas: codificação, introdução de desinformação, e recordação.

No estudo original, na fase de codificação, os participantes assistiam a um vídeo sobre um acidente de carro e eram, de seguida, divididos em diferentes grupos. A cada grupo era perguntada a velocidade a que seguia o carro no momento do acidente, com uma pequena variante: a formulação da pergunta variava de modo que cada grupo ouvia a pergunta com um verbo diferente (Loftus & Palmer, 1974). Por exemplo, a um dos grupos era colocada a questão: “A que velocidade circulavam os veículos quando se esmagaram (*smashed*) um contra o outro?” e a outro era perguntado: “A que velocidade circulavam os veículos quando colidiram (*hit*) um contra o outro?”. A principal diferença entre as duas perguntas prende-se com a intensidade que cada verbo representa. Assim, os participantes a quem tinha sido colocada a questão com o

verbo “esmagar” estimaram que os carros circulavam a uma velocidade superior, comparativamente com o grupo que respondeu à questão que continha o verbo “colidir”.

Uma semana depois, foi pedido aos mesmos participantes que descrevessem o acidente a que tinham assistido no vídeo e estes foram também questionados sobre a presença de vidros partidos (dos carros) no chão. Observou-se, que os participantes a quem tinha sido colocada a questão com o verbo “esmagar” relataram com mais frequência a presença de vidros no chão – associada a acidentes de viação mais violentos – (ainda que no vídeo não aparecesse qualquer fragmento de vidro partido), do que os grupos a quem tinha sido colocada a questão inicial com outros verbos (i.e., bater, colidir, esbarrar) (Loftus & Palmer, 1974). Estes resultados sugerem que as perguntas colocadas após um acontecimento, podem distorcer a forma como esse mesmo acontecimento é recordado - efeito de desinformação (Loftus & Palmer, 1974).

Ao longo dos anos este estudo tem sido frequentemente replicado e, ainda que com diferentes metodologias, os resultados têm sido idênticos. Na primeira fase destes estudos – fase de codificação – têm sido apresentados aos participantes diferentes tipos de informação: um vídeo de um assalto a um banco (Saraiva & Garrido, 2024; Wright et al., 2001), um assalto a uma loja de conveniência (Aizpurua et al., 2009), ou uma sequência de slides a retratar um conflito entre um professor e um aluno numa sala de aula (Zaragoza & Lane, 1994).

Na fase de introdução de informação falsa, os participantes são expostos a informações enganosas, recorrendo a diferentes metodologias, nomeadamente: participando em discussões com outras pessoas (comparsas dos investigadores) que lhes apresentam informações falsas sobre o sucedido (Wright et al., 2001), ou recebendo descrições incorretas do acontecimento, contendo detalhes e relatos de objetos que não faziam parte do cenário original (Zaragoza & Lane, 1994). Ao ser pedido aos participantes para recordarem os eventos originais (fase de recordação), tem sido evidente que os relatos destes são influenciados pelas informações introduzidas *a posteriori* (Loftus et al., 1978).

Uma das justificações apontadas para estes resultados tem por base o facto de as memórias dos acontecimentos serem construídas (codificadas) ao longo de dois momentos distintos. O primeiro remete para a informação obtida aquando do evento (e.g., a visualização do vídeo) e o segundo para a informação obtida após o evento. Quando se dá a codificação de um acontecimento, ambas as informações (corretas e falsas) são integradas na memória, pelo que se assiste a uma reconstrução mnésica do evento anteriormente codificado (Hyman & Pentland, 1996; Johnson et al., 1993; Loftus & Palmer, 1974; Loftus & Zanni, 1975).

Estes resultados mostram também que a memória é uma reconstrução ativa (e não passiva) dos acontecimentos (Loftus & Palmer, 1974; Zaragoza & Lane, 1994) e que pode ser facilmente

manipulada com a introdução de informações incorretas (desinformação) após a codificação inicial dos eventos (Loftus, 2005).

1.4. Paradigma do efeito da ilusão de verdade (e ilusão de falsidade)

Um outro paradigma a considerar nos estudos sobre desinformação é o paradigma do efeito da ilusão de verdade, que aborda os efeitos da repetição da informação na percepção da sua veracidade (Brashier & Marsh, 2020; Fazio et al., 2015). Este paradigma explora um fenómeno cognitivo no qual as pessoas tendem a considerar mais verdadeiras informações repetidas, comparativamente com informações novas (Bacon, 1979; Brashier & Marsh 2020; Hasher et al., 1977).

No estudo original (Hasher et al., 1977) foi sugerido que a repetição de algumas afirmações aumentaria a sua probabilidade de serem julgadas como verdadeiras pelos participantes (independentemente da sua veracidade factual). Para tal, os autores começaram por apresentar a uma amostra de 40 estudantes universitários um conjunto de 60 afirmações (verdadeiras e falsas) sobre diversos temas (e.g., geografia, biologia, arte, política) ao longo de três sessões. É de salientar que, no decorrer destas sessões, algumas frases foram apresentadas repetidamente, enquanto outras eram apresentadas uma única vez.

Finalmente, foi pedido aos participantes que avaliassem a veracidade de cada frase (julgamentos de verdade). Com base nos resultados obtidos nesta tarefa, a hipótese inicial colocada pelos autores confirmou-se. Isto é, os resultados mostraram que as afirmações repetidas (tanto as verdadeiras como as falsas) eram julgadas como mais verdadeiras do que as que tinham sido apresentadas apenas uma vez (Hasher et al., 1977). Por outras palavras, este estudo veio mostrar o quanto a repetição de informação aumenta a sensação de familiaridade, favorecendo a veracidade percebida (Fazio et al., 2015; Unkelbach & Rom, 2017).

Este paradigma foi posteriormente replicado com afirmações sobre diferentes temas e em diferentes contextos, nomeadamente em estudos sobre a partilha e aceitação de desinformação (De keersmaecker & Roets, 2017; Pennycook et al., 2018; Polage, 2012).

Foi evidenciado, por exemplo, que a repetição de informações falsas disfarçadas de notícias, pode levar não só à aceitação das informações como sendo verdadeiras, mas também à criação de falsas memórias (Polage, 2012) – isto é, recordações distorcidas de eventos ou até memórias de eventos que nunca ocorreram (Atkins et al., 2023; Bartlett & Burt, 1933). Por outro lado, também se observou que o efeito da ilusão de verdade pode ocorrer, mesmo em casos em que os participantes não se recordam da fonte que partilhou a informação (Begg et al., 1992).

Ainda que não tão documentado, o conceito de ilusão de falsidade é visto como sendo o efeito oposto ao da ilusão de verdade e foi observado pela primeira vez num dos estudos referenciados anteriormente (Hasher et al., 1977).

Este efeito refere-se à tendência das pessoas para julgar informações novas ou menos familiares como falsas e vem corroborar a ideia de que, ao contrário da repetição que tende a aumentar a perceção de verdade (ilusão de verdade), a novidade da informação pode levar a que a mesma seja julgada como falsa (ilusão de falsidade) (Hasher et al., 1977; Silva et al., 2017). Este resultado vem mostrar também que a familiaridade atua como uma heurística no que toca à avaliação da veracidade da informação (Metcalf et al., 1993), o que pode ser perigoso nos casos em que a informação nova até está factualmente correta (e não é vista como tal) ou que informações enganosas continuam a ser repetidas e, consequentemente, a serem julgadas como mais verdadeiras (Pennycook et al., 2020).

1.5. Estratégias de correção da desinformação

Se por um lado se pode argumentar que não é possível impedir a partilha da desinformação, sem atentar contra a liberdade de expressão dos indivíduos (Jacobs, 2022), por outro, são cada vez mais frequentes as tentativas tendentes à sua correção (Schwarz et al., 2016; van der Linden et al., 2017).

Duas das principais estratégias de correção são baseadas em abordagens corretivas (*debunking*) e preventivas (*prebunking*). A primeira procura alterar crenças, ou seja, desacreditar e corrigir informações falsas, depois de o indivíduo já ter sido exposto à informação e envolve a refutação da mesma, fornecendo evidências corretas que visam esclarecer os factos erróneos a que os indivíduos foram expostos (Lewandowsky et al., 2012; Schwarz et al., 2016).

Nas estratégias de *debunking*, a correção da desinformação é feita após a exposição e o foco está em fornecer informações claras e factualmente corretas que refutem o conteúdo enganoso (Nyhan & Reifler, 2010). A título ilustrativo: durante a pandemia COVID-19, vários profissionais de saúde adotaram esta estratégia no combate à desinformação. Para minimizar o impacto das teorias da conspiração que surgiram à volta do vírus, alguns especialistas procuraram fornecer informações corretas e precisas que contrariassem a desinformação que circulava nas redes sociais sobre o vírus (Pennycook et al., 2020).

Por outro lado, quando falamos em *pre-debunking* (ou *prebunking*) estamos a referir-nos a uma tentativa de correção da desinformação feita antes de os indivíduos serem expostos à informação (Hameleers, 2024; Tay et al., 2021). Esta estratégia baseia-se na teoria da

inoculação (McGuire, 1961), que sugere que a exposição a uma pequena amostra de desinformação, em paralelo com explicações sobre o carácter enganoso da mesma, pode tornar os indivíduos mais resistentes à desinformação (menor probabilidade de acreditar em informações falsas) (McGuire, 1961; Roozenbeek & van der Linden, 2019; van der Linden et al., 2017).

As técnicas de *prebunking* visam, assim, preparar os indivíduos para a possibilidade de se confrontarem com informações falsas dando-lhes conhecimento correto sobre um dado tema. Ainda assim, a eficácia desta estratégia pode ser comprometida pelo facto de nem sempre ser possível prever que tipo de desinformação será divulgada (Lewandowsky & van der Linden, 2021).

Também os pré e pós-avisos (ou *accuracy prompts*) têm sido documentados na literatura enquanto formas de correção da desinformação (e.g., Pennycook et al., 2021; Lin et al., 2023; Roozenbeek et al., 2022). Estes avisos tentam chamar a atenção dos leitores para a importância da verificação da informação e para uma reflexão antes da partilha da mesma (Epstein et al., 2022). Ainda assim, para além de terem apenas resultados a curto prazo, estes avisos também não melhoram a capacidade dos indivíduos de distinguir informações verdadeiras da desinformação (Kozyreva et al., 2022), o que constitui uma das principais causas da aceitação de desinformação.

Por último, a utilização das normas sociais no combate à desinformação tem também sido aplicada em alguns contextos como forma de encorajar os indivíduos a não partilhar informações falsas (Cialdini et al., 1991). Este tipo de estratégia, baseia-se na influência do grupo, e recorre, por exemplo, a indicações como “a maioria das pessoas verificou esta informação antes de partilhar” (Muddiman & Stroud, 2017), de modo que os utilizadores associem a partilha de desinformação a algo errado que devem evitar.

Apesar destas diferentes estratégias serem frequentemente utilizadas, importa destacar que não são infalíveis e, apresentam, na verdade, alguns riscos nomeadamente a possibilidade de, ao invés de corrigirem a informação falsa, reforçarem e favorecerem a mesma – *backfire effect*. Em alguns contextos e, em particular em indivíduos com fortes convicções ideológicas, a correção de desinformação pode ser interpretada como uma ameaça à sua própria identidade e valores e, como tal, aumentar a crença em informações incorretas (Nyhan & Reifler, 2010).

1.6. Objetivos e hipóteses

Este estudo tem como objetivo investigar o papel da credibilidade da fonte na aceitação e correção de desinformação em contextos de saúde.

Na primeira fase do procedimento – fase de exposição – os participantes veem um conjunto de frases falsas relacionadas com saúde (de uma fonte credível, não credível, ou sem fonte) e fazem um julgamento de verdade para cada uma. Na segunda fase – introdução de (des)informação – é apresentado um novo conjunto de afirmações. Algumas destas serão repetições das anteriormente apresentadas, outras serão contradições (i.e., versão verdadeira das afirmações anteriormente apresentadas), e outras completamente novas. Para cada uma destas afirmações, os participantes realizam um novo julgamento de verdade seguindo o procedimento da fase anterior.

Colocaram-se as seguintes hipóteses:

Na fase de exposição, os julgamentos de verdade serão mais elevados quando a afirmação foi apresentada por uma fonte credível (vs. fonte não credível) – H1.

Na fase de reexposição, os julgamentos de verdade para afirmações repetidas serão mais elevados (do que para afirmações novas) replicando o efeito de ilusão de verdade – H2a, enquanto as contradições serão consideradas mais falsas (do que as afirmações novas) – efeito da ilusão de falsidade – H2b.

Esperam-se também que as repetições da informação inicialmente apresentada por uma fonte credível (vs. não credível) sejam consideradas mais verdadeiras – H3.

CAPÍTULO 2.

Estudo Piloto

O objetivo deste estudo foi desenvolver e selecionar um conjunto de perfis de utilizadores de redes sociais com base na sua elevada ou baixa credibilidade. Para tal, criámos perfis com diferentes profissões e fotografias. Este estudo foi aprovado pela Comissão de ética do Ispa-Instituto Universitário (Ref: I-097-12-22).

2.1. Método

Participantes

Neste pré-teste participaram 51 indivíduos, dos quais quatro foram excluídos por terem falhado os *attention checks*. Assim, a amostra final incluiu 47 participantes (72% do sexo feminino), com idades compreendidas entre os 19 e os 72 anos ($M = 30.02$; $DP = 14.08$).

Em termos de utilização de redes sociais, o Instagram (90.20%) e o WhatsApp (94.12%) foram referidos pela maioria, seguidos do Facebook (50.98%), LinkedIn (43.14%), TikTok (37.25%) e X (antigo Twitter) (27.45%).

Material

O material utilizado no pré-teste consistiu em 32 imagens de perfis de Instagram e LinkedIn criados para este estudo. A profissão escolhida e a biografia foram criadas de forma a transmitir elevada ou baixa credibilidade e a fotografia de perfil foi recolhida na internet. Os perfis foram criados com recurso ao Mediamodifier, Canva e Fake Detail Generator.

Os perfis foram organizados em duas categorias: credíveis e não credíveis, sendo que cada categoria contava com 16 perfis (4 femininos no Instagram, 4 masculinos no Instagram, 4 femininos no LinkedIn e 4 masculinos no LinkedIn). Os perfis criados com o intuito de se assemelharem a fontes credíveis diziam respeito a professores universitários, médicos internistas, investigadores em ciências médicas e especialistas em saúde pública, enquanto os perfis correspondentes a fontes não credíveis eram *life coaches*, *personal trainers*, criadores de conteúdos digitais (i.e., *influencers*) e terapeutas holísticos.

Cada perfil (e.g., médico internista) tinha uma versão masculina e feminina, em cada uma das redes sociais utilizadas (e.g., Médico internista x Instagram; Médica internista x Instagram; Médico internista x LinkedIn; Médica Internista x LinkedIn).

Procedimento

O pré-teste foi programado na plataforma Qualtrics e posteriormente disponibilizado através de um link, divulgado em algumas redes sociais (e.g., WhatsApp, Instagram). Uma parte dos dados

foi recolhida em laboratório, tendo os participantes sido compensados com créditos pela sua participação.

Todos os participantes foram devidamente informados acerca do âmbito do estudo, da sua duração aproximada (25 minutos), do carácter voluntário da sua participação e do anonimato dos dados recolhidos. Após a aceitação destes termos, foram recolhidas informações sobre a idade e género, assim como as redes sociais utilizadas pelo participante.

De seguida, foram apresentadas instruções relativamente à tarefa que consistia em avaliar cada um dos 32 perfis num conjunto de dimensões: Sociabilidade (Amável, Simpático(a), Caloroso(a), Sociável); Competência (Inteligente, Competente, Criativo(a), Eficiente), Dominância (Prestigiado(a), Poderoso(a), Influyente, Importante); Confiabilidade (Responsável, Consciencioso(a), Cuidadoso(a), Prudente); e Credibilidade (Respeitável, Confiável, Perito(a), Credível). Para tal, utilizaram uma escala de 7 pontos (1 = *Nada* a 7 = *Muito*).

Todos os participantes viam cada um dos 32 perfis, de forma aleatória e sem tempo limite para fazerem a avaliação dos mesmos (ver Anexos).

2.2. Resultados

Após obtidas as médias para cada dimensão e para cada perfil, estes valores foram ordenados de forma decrescente, de modo que os valores mais elevados correspondessem a perfis mais credíveis e os valores mais baixos a perfis menos credíveis.

Após esta ordenação, procurou-se encontrar, para a condição credível, um exemplar de cada perfil (i.e., Feminino x LinkedIn, Feminino x Instagram, Masculino x LinkedIn e Masculino x Instagram) que também fosse classificado com valores elevados nas restantes dimensões (e.g., confiabilidade). Um procedimento semelhante foi realizado para seleccionar os perfis pouco credíveis, e seleccionados os perfis com valores mais baixos nas restantes dimensões.

Esta análise (ver Tabela 2.1.) permitiu seleccionar os seguintes perfis para o estudo principal:

- Credíveis: médico internista (perfil masculino) e investigadora em ciências médicas (perfil feminino), nas suas versões do LinkedIn e Instagram.
- Não credíveis: terapeuta holístico (perfil masculino) e criadora de conteúdos (perfil feminino), nas suas versões do LinkedIn e Instagram.

Tabela 2.1. *Avaliação dos perfis selecionados*

Perfil	Dimensões avaliativas				
	Credibilidade	Confiança	Competência	Dominância	Sociabilidade
	<i>M(DP)</i>	<i>M(DP)</i>	<i>M(DP)</i>	<i>M(DP)</i>	<i>M(DP)</i>
Fonte credível					
Médico Instagram	5.46 (1.40)	5.19 (1.36)	5.12 (1.39)	5.16 (1.50)	4.67 (1.33)
Médico LinkedIn	5.45 (1.24)	5.16 (1.24)	5.18 (1.23)	5.17 (1.32)	4.63 (1.26)
Investigadora Instagram	5.45 (1.36)	5.06 (1.39)	5.20 (1.43)	5.05 (1.40)	4.68 (1.35)
Investigadora LinkedIn	5.57 (1.21)	5.19 (1.26)	5.23 (1.24)	5.10 (1.23)	4.68 (1.26)
Fonte não credível					
Terapeuta holístico Instagram	3.50 (1.41)	3.88 (1.27)	3.72 (1.35)	3.43 (1.27)	4.16 (1.33)
Terapeuta holístico LinkedIn	3.39 (1.20)	3.66 (1.21)	3.61 (1.18)	3.37 (1.14)	4.26 (1.26)
Criadora conteúdos Instagram	3.04 (1.18)	3.34 (1.22)	3.59 (1.24)	3.28 (1.24)	3.89 (1.42)
Criadora conteúdos LinkedIn	3.59 (1.31)	3.76 (1.20)	3.99 (1.26)	3.45 (1.09)	4.56 (1.27)

CAPÍTULO 3.

Estudo Experimental

Neste estudo experimental procurou-se investigar de que forma a credibilidade da fonte influencia os julgamentos de veracidade de informações sobre saúde divulgadas em duas redes sociais.

3.1. Método

Participantes

Uma amostra de 90 participantes foi determinada por uma análise de poder realizada à priori (G*Power) usando como referência um efeito médio a grande ($n_p^2 = .10$) e um poder $1-\beta = 0.80$ para detetar diferenças nos julgamentos de verdade na fase de reexposição entre frases repetidas e novas (Efeito de Ilusão de Verdade) e entre contradições e frases novas (Efeito de Ilusão de Falsidade). A amostra final consistiu em 100 participantes, dos quais 44 eram mulheres, 54 eram homens e 2 identificaram-se como “outro”, com idades compreendidas entre os 20 e os 52 anos ($M = 29.26$, $DP = 7.37$).

Relativamente às suas habilitações literárias, 20 dos participantes indicaram ter concluído o Ensino Secundário/12º ano e 80 o Ensino Superior (e.g., Licenciatura ou outro grau académico).

Os participantes foram recrutados através da plataforma Prolific Academic e recompensados com £2.

Design

O design experimental deste estudo era do tipo fatorial misto. O tipo de frase apresentada na fase de reexposição foi manipulado intra-participantes (repetida vs. nova vs. contradição), assim como a veracidade das frases (verdadeira vs. falsa). O tipo de fonte (credível vs. não credível vs. sem fonte) foi manipulado entre participantes.

Material

Neste estudo foram utilizados os 8 perfis seleccionados do estudo piloto e foram criados 2 perfis adicionais sem fonte (do LinkedIn e Instagram) respeitantes a três categorias – fontes credíveis, fontes não credíveis e sem fonte (ver Anexos).

Para a primeira fase do estudo foram utilizadas 30 frases falsas (e.g., O leite é a maior fonte de cálcio) sobre saúde. Para a segunda fase, foram utilizadas 10 frases repetidas (falsas), 10

contradições (verdadeiras), e 10 frases (cinco verdadeiras e cinco falsas) novas. Estas frases foram selecionadas de Santos (2023).

Para garantir um maior controlo experimental, foram criadas três versões dos estímulos a apresentar na segunda fase do estudo (ver Anexos). Estas versões permitiam que todas as frases apresentadas inicialmente fossem repetidas (10 em cada versão) e contraditas (10 em cada versão), numa manipulação entre sujeitos (nv1 = 32 participantes, nv2 = 33 participantes, nv3 = 35 participantes).

Tabela 3.1.

Média e desvio padrão da avaliação de veracidade das frases utilizadas no questionário (por cada tipo de frase e versão)

Tipo de Frase	Versão					
	1		2		3	
	<i>M</i>	<i>DP</i>	<i>M</i>	<i>DP</i>	<i>M</i>	<i>DP</i>
Repetições	2.97	0.15	2.88	0.11	2.72	0.13
Contradições	3.72	0.11	3.71	0.14	3.94	0.09
Novas	3.33	0.07	3.25	0.07	3.03	0.07

Procedimento

Este estudo foi aprovado pela Comissão de Ética do Ispa-Instituto Universitário (Ref: I-097-12-22). O estudo foi programado no Qualtrics. O questionário iniciava-se com a leitura e aceitação de um consentimento informado, no qual os participantes tomavam conhecimento do carácter voluntário do estudo e da possibilidade de desistir a qualquer momento, se assim o entendessem. Nesta fase, os participantes também eram informados de que deveriam ter 18 ou mais anos de idade para participar no estudo e que a informação recolhida durante o mesmo era anónima e confidencial.

Após a aceitação deste termo de consentimento, foram colocadas algumas questões sociodemográficas aos participantes. Neste sentido, foi-lhes pedido que indicassem a sua idade (em anos), o seu género e as suas habilitações literárias.

De seguida, os participantes foram informados de que lhes seriam apresentadas algumas informações sobre saúde e bem-estar partilhadas por diferentes profissionais, nas suas redes sociais. Eram ainda informados que deveriam prestar o máximo de atenção possível a cada um dos perfis e às informações partilhadas, pois mais tarde teriam de recordar parte dessa

informação. Os participantes eram ainda informados que, à semelhança do que acontece nas redes sociais, algumas dessas informações eram falsas e outras eram verdadeiras. Esta tarefa inicial consistia em ler cada perfil, assim como as informações apresentadas e avaliá-las quanto à sua veracidade, numa escala de 1 (completamente falsa) a 6 (completamente verdadeira). Os participantes eram alocados a uma das três condições experimentais (fonte credível, fonte não credível ou sem fonte) de forma aleatória.

Na fase de exposição, os participantes viam, inicialmente, apenas a imagem de um dos perfis mencionados anteriormente durante 15 segundos (Figura 3.1.); depois eram apresentadas de forma aleatória, 30 publicações feitas por esse mesmo perfil e que continham, cada uma, uma frase sobre saúde/bem-estar (Figura 3.2.). Para cada uma destas publicações/frases, os participantes tinham de fazer um julgamento de verdade na escala referida.



Figura 3.1. Exemplo de perfil apresentado aos participantes



Figura 3.2. Exemplo de publicação associada a um dos perfis

Em seguida, os participantes eram questionados acerca do sexo e profissão da pessoa cujo perfil tinham acabado de ver (*attention check*).

Seguia-se uma tarefa distratora, que consistia na apresentação aleatória de dígitos, entre 1 e 9. A tarefa dos participantes consistia em identificar se cada dígito era par ou ímpar, escolhendo a opção correspondente. Esta tarefa tinha a duração de 1 minuto.

De seguida, os participantes eram informados que veriam um novo conjunto de informações sobre saúde e bem-estar, sendo que algumas dessas informações eram verdadeiras e outras falsas. Tal como na fase de exposição, os participantes avaliavam a veracidade dos *posts* numa escala de 1 (completamente falsa) a 6 (completamente verdadeira).

Nesta fase, os participantes viam novamente 30 frases, sem referência a qualquer perfil ou fonte de informação. Um terço das frases eram repetidas (ou seja, já tinham sido apresentadas na primeira tarefa), um terço eram contradições de frases que tinham aparecido na fase de exposição, e as restantes eram informações novas. Desta vez, as frases não eram apresentadas sob a forma de um *post* de uma rede social.

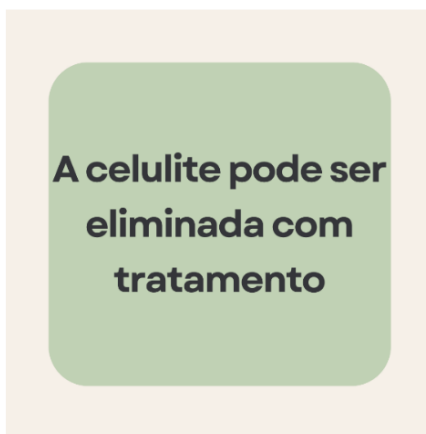


Figura 3.3. Exemplo de uma das frases (repetição) apresentadas na fase de reexposição

Concluída esta fase, os participantes completaram novamente uma tarefa distratora de classificação de dígitos idêntica à primeira (1 minuto de duração).

No final do estudo os participantes avaliavam o perfil nas seguintes dimensões: competência, confiança, calorosidade e influência, numa escala de 7 pontos (1 = pouco a 7 = muito).

3.2. Resultados

Avaliação do perfil

A avaliação das quatro dimensões de cada perfil (i.e., caloroso, credível, confiável e influente) foi agregada num índice geral com elevada consistência interna ($\alpha = 0.86$) (Cronbach, 1951).

A avaliação deste índice em função da fonte (credível vs. não-credível vs. sem fonte) foi analisada através de uma *One-Way* ANOVA e revelou uma diferença estatisticamente significativa entre as três fontes na avaliação do perfil, $F(2,97) = 22.93, p < .001, \eta_p^2 = .321$. As comparações *pairwise* revelaram que os participantes na condição de fonte credível deram avaliações mais elevadas ao perfil ($M = 3.56, DP = .86$), do que os da fonte não-credível ($M = 2.92, DP = 1.20$), $p = .025$, e do que os da condição sem fonte ($M = 1.94, DP = .80$), $p < .001$. Os participantes da condição não-credível também avaliaram melhor o perfil do que os da condição sem fonte, $p < .001$.

Efeito da fonte nos julgamentos de verdade na fase de exposição

Uma *One-Way* ANOVA usando a fonte como fator (credível vs. não-credível vs. sem fonte), revelou uma diferença estatisticamente significativa entre as três fontes no julgamento de verdade das frases, $F(2,97) = 10.28, p < .001, \eta_p^2 = .175$. Comparações *pairwise* revelaram que as avaliações de veracidade foram mais baixas na condição sem fonte ($M = 2.28, DP = .60$), do que na condição fonte credível ($M = 2.99, DP = .59; p < .001$), e fonte não-credível ($M = 2.70, DP = .72; p = .031$). Ao contrário do esperado, não existiram diferenças significativas entre as fontes credível e não-credível, $p = .185$, na avaliação que os participantes fizeram da veracidade das frases apresentadas.

Efeito de ilusão de verdade

Para estudar o efeito de ilusão de verdade, comparou-se a avaliação de veracidade realizada na fase de reexposição das frases repetidas (falsas) com as novas-falsas, através de um teste *t* para amostras emparelhadas. Ao contrário do esperado, não se verificou o efeito de ilusão de verdade, $t(99) = -0.30, p = .765$, na medida em que não se observaram diferenças significativas entre a avaliação de veracidade das frases repetidas ($M = 2.85, DP = .77$) e das frases novas-falsas ($M = 2.87, DP = .77$).

Efeito de ilusão de falsidade

Recorreu-se, igualmente, a um teste *t* para analisar o efeito de ilusão de falsidade, que procurou comparar, na fase de reexposição, a avaliação de veracidade das contradições (verdadeiras) e das frases novas-verdadeiras. Os resultados revelaram um efeito significativo na direção oposta

à da hipótese colocada inicialmente, $t(99) = 3.40, p = .001$. Quer isto dizer que a avaliação de veracidade das contradições ($M = 3.80, DP = .69$) foi significativamente superior à das frases novas-verdadeiras ($M = 3.53, DP = .73$).

Análises adicionais

Os efeitos de ilusão de verdade e falsidade podem também ser explorados, comparando a avaliação de veracidade das frases repetidas apresentadas na fase de reexposição com as frases apresentadas inicialmente e a avaliação de veracidade das contradições apresentadas na fase de reexposição com as frases apresentadas inicialmente, respetivamente.

Assim, comparámos as 10 frases repetidas na fase de reexposição com as 10 frases correspondentes apresentadas inicialmente, através de um teste t para amostras emparelhadas. Este teste, evidenciou o efeito de ilusão de verdade, $t(99) = 6.20, p < .001$, no qual avaliação de veracidade das frases repetidas ($M = 2.85, DP = .77$) foi significativamente superior à das frases apresentadas inicialmente ($M = 2.66, DP = .77$).

Testou-se também novamente o efeito de ilusão de falsidade, através da comparação da avaliação de veracidade das 10 contradições apresentadas na fase de reexposição com as 10 frases correspondentes apresentadas inicialmente. Aqui, verificou-se um efeito significativo na direção oposta à hipotetizada, $t(99) = 8.24, p < .001$. Ou seja, a avaliação de veracidade das contradições ($M = 3.80, DP = .69$) foi significativamente superior à das frases apresentadas inicialmente [e que, efetivamente, eram falsas] ($M = 2.68, DP = .80$).

Efeito da credibilidade da fonte nos julgamentos de veracidade na fase de reexposição

Finalmente, analisámos o efeito da fonte nas avaliações de veracidade realizadas durante a fase de reexposição. Uma vez que nesta fase os perfis não eram apresentados aos participantes, apenas foram consideradas as avaliações de veracidade das frases repetidas, inicialmente apresentadas associadas a uma fonte específica. Uma *One-Way* ANOVA revelou diferenças estatisticamente significativa na avaliação de veracidade das frases repetidas em função da fonte, $F(2, 97) = 5.56, p = .005, \eta_p^2 = .103$. As comparações *pairwise* mostraram que os participantes da condição fonte credível avaliaram as frases repetidas como mais verdadeiras ($M = 3.14, DP = .62$), do que os da condição sem fonte ($M = 2.54, DP = .72$), $p = .004$. As comparações entre as restantes fontes não foram significativas: credível vs. não-credível ($Mdif = .31; DP = .18$), $p = .260$ e sem fonte vs. não-credível ($Mdif = -0.30; DP = .19$), $p = .343$.

Ao comparar apenas os participantes das condições credível e não-credível, a *One-Way* ANOVA os resultados, ainda que não significativos apresentam o padrão esperado, $F(2,67) = 2.94, p = .091, \eta_p^2 = 0.042$. Os participantes na condição fonte credível ($M = 3.14, DP = .62$)

avaliaram as frases repetidas como mais verdadeiras do que os da condição fonte não-credível ($M = 2.83$, $DP = .86$).

Attention checks

Os participantes identificaram corretamente o género do perfil em 94% ($DP = 0.24$) das vezes e a profissão em 78% ($DP = 0.42$). Estas elevadas taxas de acerto sugerem que os participantes prestaram atenção à informação contida no perfil durante a sua apresentação.

CAPÍTULO 4.

Discussão

A desinformação online é considerada um dos maiores desafios da era digital, com impactos significativos na política, na saúde pública e na percepção social dos acontecimentos (Delmazo, 2018; Galhardi et al., 2020). Existem já inúmeros estudos que debatem a propagação da desinformação na área da saúde (Ghenai & Mejova, 2018) e que procuram explicar os fenómenos subjacentes à sua aceitação e transmissão nas redes sociais (Scherer et al., 2021; Shao et al., 2018).

Com isto, têm surgido diferentes tentativas de correção da desinformação, numa ótica de combater (ou pelo menos abrandar) a sua propagação (Shu et al., 2020). Uma dessas tentativas prende-se com a manipulação da credibilidade da fonte que partilha a informação. Isto porque os utilizadores tendem a confiar mais em informações provenientes de fontes que consideram legítimas e especialistas no assunto em causa (Ecker et al., 2020). Aliado a isso, vários estudos têm demonstrado que a correção de desinformação é mais eficaz quando acompanhada de referências a fontes credíveis (Vraga & Bode, 2017). Ainda assim, não são garantia de sucesso (Epstein et al., 2022; Roozenbeek & van der Linden, 2022).

O presente trabalho procurou investigar o papel da credibilidade da fonte na aceitação e correção de desinformação em contextos de saúde, baseado nos pressupostos do paradigma da desinformação (Loftus & Palmer, 1974) e do efeito de ilusão de verdade (Hasher et al., 1977) e ilusão de falsidade (Pennycook et., 2020).

4.1. Principais resultados

Numa primeira fase, foi conduzido um estudo piloto com o objetivo de determinar que tipo de perfis eram considerados credíveis e não credíveis pelos participantes. Estes perfis foram avaliados de acordo com um conjunto de dimensões (sociabilidade, competência, dominância, confiabilidade e credibilidade). Conforme seria de esperar, os profissionais ligados à saúde, ciência e educação foram considerados mais credíveis, o que já vem sendo evidenciado noutros estudos (Fiske et al., 2002; Ipsos, 2022).

Por outro lado, apesar de ainda não existirem tantos estudos que recorram especificamente a perfis de *influencers* e terapeutas holísticos para estudar a desinformação online em contexto de saúde, já existem algumas evidências de que os *influencers* são considerados menos credíveis que outros profissionais. De facto, estes são por vezes percebidos como menos úteis, trabalhadores, e credíveis que profissões mais tradicionais (Yilmazdoğan al., 2021) e são

também considerados menos credíveis no que toca à partilha de informação sobre saúde (Sabbagh et al., 2020) e mais propícios a espalhar informações incorretas (Hendry et al., 2022).

Os resultados do estudo principal revelaram, em primeiro lugar, que a credibilidade da fonte teve um impacto significativo nos julgamentos de verdade dos participantes. As frases partilhadas por fontes consideradas credíveis foram descritivamente consideradas mais verdadeiras do que as divulgadas por fontes não credíveis. Estes resultados estão alinhados com estudos anteriores que mostram que credibilidade da fonte é um fator importante na aceitação da desinformação, na medida em que os indivíduos tendem a confiar mais em conteúdos partilhados por especialistas ou instituições/organizações reconhecidas (Ecker et al., 2020; Roozenbeek & van der Linden, 2022). No entanto, esta interpretação não pode ser diretamente aplicada aos nossos resultados atendendo a que a diferença nos julgamentos de veracidade de frases partilhadas por fontes credíveis e não credíveis não atingiu significância estatística. Desta forma não foi totalmente confirmada a nossa Hipótese 1. Observou-se ainda que a ausência de fonte levou a avaliações de veracidade mais baixas do que as das frases partilhadas por fontes credíveis e não credíveis. Este resultado pode ser explicado com base no facto de que, nas redes sociais, a incerteza sobre a origem da informação pode gerar desconfiança - em vez de uma suposta neutralidade ou indiferença – (Hasnat, 2014).

No que se refere ao efeito da credibilidade da fonte na avaliação das frases repetidas na fase de reexposição observou-se um padrão semelhante. Descritivamente, as frases repetidas inicialmente apresentadas por fontes credíveis foram avaliadas como mais verdadeiras do que as apresentadas por fontes não credíveis e sem fonte. No entanto, apenas a diferença entre fontes credíveis e sem fonte foi significativa. Ainda assim, este padrão apoiou parcialmente a nossa Hipótese 3.

Relativamente ao efeito da ilusão de verdade, este estudo não encontrou diferenças significativas entre a avaliação de veracidade de frases repetidas (falsas) e a avaliação das frases novas (falsas). Este resultado difere dos estudos clássicos nesta área (e.g., Begg et al., 1992; Hasher et al., 1977; Unkelbach & Rom, 2017), nos quais a repetição aumenta a percepção de veracidade da informação. Isto pode dever-se ao facto de, ao contrário da maioria dos estudos, todas as frases utilizadas na fase de exposição serem falsas. Tal pode ter levado os participantes a suspeitar que as frases eram na generalidade todas falsas com consequências na variabilidade dos seus julgamentos de veracidade subsequentes. Por outro lado, uma vez que o tema das frases – saúde e bem-estar – pode ser do conhecimento dos participantes, a sua capacidade para reconhecerem a falsidade de toda informação pode também ter influenciado os seus julgamentos de veracidade. Embora existam estudos que mostram o efeito de ilusão de verdade

quando os participantes têm conhecimento prévio acerca das informações apresentadas, incluindo informação sobre saúde (Vellani et al., 2023), a maioria das investigações prévias sugerem que o efeito de ilusão de verdade é menos acentuado quando os participantes têm um elevado grau de certeza acerca da falsidade da informação (Fazio et al., 2015; Pennycook & Rand, 2018).

Verificou-se ainda um efeito de ilusão de falsidade na direção oposta à esperada. Quer isto dizer que as frases que contradiziam informações (falsas) anteriormente vistas foram avaliadas como significativamente mais verdadeiras do que as frases novas (verdadeiras). Este efeito pode dever-se ao facto de, mais uma vez, os participantes terem inferido que a informação inicialmente apresentada era falsa. Assim, ao serem confrontados com contradições (verdadeiras), podem considerá-las como uma correção da informação apresentada inicialmente (Unkelbach & Rom, 2017). Por este motivo reconhecem nas contradições informação verdadeira, um resultado que não deixa de ser interessante.

Quando se avalia a ilusão de verdade e falsidade com base na comparação direta entre os julgamentos de verdade na fase de exposição e reexposição observa-se o efeito de ilusão de verdade, mas o de falsidade mantêm-se invertido. Estes resultados mostram que a repetição aumenta a perceção de veracidade apoiando parcialmente a nossa Hipótese 2a. No entanto, as contradições, foram avaliadas como mais verdadeiras na fase de reexposição (contrariando a nossa Hipótese 2b), provavelmente pelo reconhecimento por parte dos participantes da sua natureza verdadeira, tal como exposto anteriormente.

4.2. Limitações e sugestões para estudos futuros

Numa ótica de reflexão final, foram identificadas algumas limitações deste estudo. A primeira delas prende-se com a operacionalização da condição sem fonte enquanto condição de credibilidade intermédia (entre a condição credível e a não credível). Esta condição, que foi pensada como uma fonte anónima e neutra, acabou por ser julgada pelos participantes como menos credível que as restantes. De facto, a ausência de fotografia e informação no perfil poderá ter despertado alguma desconfiança nos participantes. Num estudo futuro, que pretenda manter uma condição de controlo, recomenda-se a utilização de uma fotografia de perfil, acompanhada de alguma informação não relacionada com saúde.

Por outro lado, identificámos como problemática a utilização exclusiva de frases falsas na fase de exposição. Ao contrário da maioria dos estudos anteriores, que apresentam informação verdadeira e falsa, a apresentação exclusiva de informação falsa, poderá ter extinguido o efeito de ilusão de verdade, quando determinado a partir da comparação de frases falsas repetidas com

frases novas. Além disso, embora relativamente ambíguas quanto à sua veracidade, algumas das frases podem ter sido reconhecidos como falsas por alguns participantes, uma vez que ao contrário da maioria dos estudos anteriores, não se tratava de informações triviais, mas de frases sobre saúde sobre as quais poderiam ter algum conhecimento prévio. Ainda assim, quando comparada a veracidade na fase de exposição e de repetição observou-se o efeito de ilusão de verdade, sugerindo que a repetição aumenta a veracidade percebida. Num estudo futuro, sugere-se a utilização de frases verdadeiras e falsas na fase de exposição como forma de atenuar este padrão de resposta. A utilização de informação mais ambígua e a utilização de frases verdadeiras e falsas poderão também contribuir para atenuar julgamentos de veracidade elevada generalizados, para as contradições apresentadas na fase de reexposição.

Conclusão

Os resultados deste estudo sugerem que a credibilidade da fonte pode desempenhar um papel importante na forma como a veracidade da (des)informação é julgada, particularmente em contextos de saúde. A ausência de fonte gerou maior desconfiança, enquanto fontes credíveis tendem a contribuir para uma maior aceitação da informação, mesmo quando esta é factualmente falsa (Ecker et al., 2020; Metzger & Flanagin, 2013). Embora nem todas as hipóteses tenham sido confirmadas, os dados reforçam a necessidade de considerar a origem da informação na luta contra a desinformação, sobretudo em áreas críticas como a saúde pública. Tal como demonstrado por Pennycook e colegas (2020), a familiaridade e a confiança percebida na fonte podem sobrepor-se à veracidade da informação, o que reforça a importância da promoção da literacia e da verificação das fontes no combate à desinformação.

Referências Bibliográficas

- Ahmed, S., & Tan, H. W. (2022). Personality and perspicacity: Role of personality traits and cognitive ability in political misinformation discernment and sharing behavior. *Personality and Individual Differences*, 196, 111747. <https://doi.org/10.1016/j.paid.2022.111747>
- Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2009). False memories for a robbery in young and older adults. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(2), 174-187. <https://doi.org/10.1002/acp.1461>
- Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election* (NBER Working Paper No. 23089). National Bureau of Economic Research. <https://doi.org/10.3386/w23089>
- Atkins, C., Zhao, B. Z. H., Asghar, H. J., Wood, I., & Kaafar, M. A. (2023). Those aren't your memories, they're somebody else's: seeding misinformation in chat bot memories. In D. Pointcheval & D. Vergnaud (Eds.), *Applied cryptography and network security* (Vol. 13965, pp. 284–308). Springer. https://doi.org/10.1007/978-3-031-33488-7_11
- Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, 5(3), 241-252. <https://doi.org/10.1037/0278-7393.5.3.241>
- Bailey, S. C., Brega, A. G., Crutchfield, T. M., Elasy, T., Herr, H., Kaphingst, K., Karter, A. J., Moreland-Russell, S., Osborn, C. Y., Pignone, M., Rothman, R., & Schillinger, D. (2014). Update on health literacy and diabetes. *The Diabetes Educator*, 40(5), 581-604. <https://doi.org/10.1177/0145721714540220>
- Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154-175. <https://doi.org/10.1080/21670811.2017.1345645>
- Bartlett, F. C., & Burt, C. (1933). Remembering: A study in experimental and social psychology. *British Journal of Educational Psychology*, 3(2), 187–192. Portico. <https://doi.org/10.1111/j.2044-8279.1933.tb02913.x>
- Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, 121(4), 446-458. <https://doi.org/10.1037/0096-3445.121.4.446>
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192-205. <https://doi.org/10.1509/jmr.10.0353>
- Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), 97-107. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>
- Bessi, A., Zollo, F., Del Vicario, M., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Trend of narratives in the age of misinformation. *PLoS ONE*, 10(8), e0134641. <https://doi.org/10.1371/journal.pone.0134641>
- Betsch, C., Renkewitz, F., Betsch, T., & Ulshöfer, C. (2010). The influence of vaccine-critical websites on perceiving vaccination risks. *Journal of Health Psychology*, 15(3), 446-455. <https://doi.org/10.1177/1359105309353647>
- Blaskiewicz, R. (2013). The big pharma conspiracy theory. *Medical Writing*, 22(4), 259-261. <https://doi.org/10.1179/2047480613Z.000000000142>
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131-1140. <https://doi.org/10.1080/10410236.2017.1331312>

- Boom, J. A., Cunningham, R. M., & McGee, L. U. (2018). Vaccine myths: Setting the record straight. *Journal of Family Strengths*, 18(1), Article 13. <https://doi.org/10.58464/2168-670X.1375>
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7-17. <https://doi.org/10.1038/s41467-018-07761-2>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318. <https://doi.org/10.1073/pnas.1618923114>
- Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, 194, 104054. <https://doi.org/10.1016/j.cognition.2019.104054>
- Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., Ruths, D., Teichmann, L., & Zhilin, O. (2020). The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *Harvard Kennedy School Misinformation Review*, 1(3), 1-18. <https://doi.org/10.37016/mr-2020-028>
- Budak, C., Nyhan, B., Rothschild, D. M., Thorson, E., & Watts, D. J. (2024). Misunderstanding the harms of online misinformation. *Nature*, 630(8015), 45-53. <https://doi.org/10.1038/s41586-024-07417-w>
- Calvillo, D. P., Ross, B. J., Garcia, R. J. B., Smelter, T. J., & Rutchick, A. M. (2021). Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, 12(8), 1119-1128. <https://doi.org/10.1177/1948550620940539>
- Calvillo, D. P., León, A., & Rutchick, A. M. (2024). Personality and misinformation. *Current opinion in psychology*, 55, 101752. <https://doi.org/10.1016/j.copsyc.2023.101752>
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, 24, 201-234. [https://doi.org/10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653-665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Davis, T. C., Wolf, M. S., Bass, P. F., III, Thompson, J. A., Tilson, H. H., Neuberger, M., & Parker, R. M. (2006). Literacy and misunderstanding prescription drug labels. *Annals of internal medicine*, 145(12), 887-894. <https://doi.org/10.7326/0003-4819-145-12-200612190-00144>
- De keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, 65, 107-110. <https://doi.org/10.1016/j.intell.2017.10.005>
- Delmazo, C., & Valente, J. C. L. (2018). Fake news nas redes sociais online: propagação e reações à desinformação em busca de cliques. *Media & Jornalismo*, 18(32), 153-169. https://doi.org/10.14195/2183-5462_32_11
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrocioni, W. (2016a). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific Reports*, 6, 37825. <https://doi.org/10.1038/srep37825>
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrocioni, W. (2016b). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559. <https://doi.org/10.1073/pnas.1517441113>
- De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude.

- International Journal of Advertising*, 36(5), 798-828.
<https://doi.org/10.1080/02650487.2017.1348035>
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880-896.
<https://doi.org/10.1037/0022-3514.93.5.880>
- Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, 68, 1-7. <https://doi.org/10.1016/j.chb.2016.11.009>
- Ecker, U. K. H., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5(1). <https://doi.org/10.1186/s41235-020-00241-6>
- Epstein, Z., Foppiani, N., Hilgard, S., Sharma, S., Glassman, E., & Rand, D. (2022). Do explanations increase the effectiveness of AI-crowd generated fake news warnings? *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 183-193.
<https://doi.org/10.1609/icwsm.v16i1.19283>
- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993-1002.
<https://doi.org/10.1037/xge0000098>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902.
<https://doi.org/10.1037/0022-3514.82.6.878>
- Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Advances in Political Psychology*, 38(S1), 127-150. <https://doi.org/10.1111/pops.12394>
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117(1), 39-66. <https://doi.org/10.1037/0033-2909.117.1.39>
- Galhardi, C. P., Freire, N. P., Minayo, M. C. S., & Fagundes, M. C. M. (2020). Fato ou fake? Uma análise da desinformação frente à pandemia da Covid-19 no Brasil. *Ciência & Saúde Coletiva*, 25(Suppl 2), 4201-4210. <https://doi.org/10.1590/1413-812320202510.2.28922020>
- Ghenai, A., & Mejova, Y. (2018). Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1-20. <https://doi.org/10.1145/3274327>
- Gillespie, T. (2014). The relevance of algorithms. In P. Boczkowski, K. Foot, & T. Gillespie (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-193). MIT Press. <https://doi.org/10.7551/mitpress/9780262525374.003.0009>
- Godlee, F., Smith, J., & Marcovitch, H. (2011). Wakefield's article linking MMR vaccine and autism was fraudulent. *BMJ*, 342, c7452. <https://doi.org/10.1136/bmj.c7452>
- Graziano, W. G., & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of Personality*, 70(5), 695-728. <https://doi.org/10.1111/1467-6494.05021>
- Greifeneder, R., Jaffé, M. E., Newman, E. J., & Schwarz, N. (Eds.). (2020). *The psychology of fake news: Accepting, sharing, and correcting misinformation*. Routledge.
<https://doi.org/10.4324/9780429295379>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374-378.
<https://doi.org/10.1126/science.aau2706>

- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Hameleers, M., & van der Goot, E. (2024). Look at what the real facts and experts say! The use of expert references and objectivity claims in disinformation: A qualitative exploration and typology. *Journalism*, 26(7), 1469-1487. Advance online publication. <https://doi.org/10.1177/14648849241257383>
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107-112. [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1)
- Hasnat, M. O. (2014). Credibility of social online media: in the eyes of Finnish professional journalists. *Comunicação e Sociedade*, 25, 220-235. [https://doi.org/10.17231/comsoc.25\(2014\).1870](https://doi.org/10.17231/comsoc.25(2014).1870)
- Hendry, N. A., Hartung, C., & Welch, R. (2022). Health education, social media, and tensions of authenticity in the ‘influencer pedagogy’ of health influencer Ashy Bines. *Learning, Media and Technology*, 47(4), 427–439. <https://doi.org/10.1080/17439884.2021.2006691>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press. https://doi.org/10.1007/978-3-658-45292-6_33
- Hyman, I. E., Jr., & Pentland, W. E. (1996). False memories of childhood experiences. *Applied Cognitive Psychology*, 10(4), 371-381. <http://dx.doi.org/10.1002/acp.2350090302>
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ipsos (2022). *Global Trustworthiness Index 2022*. Ipsos. Recuperado de <https://www.mapfre.com/en/insights/commitment/professions-instill-trust/>
- Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A.-H. M., Hasan, S. M. M., Kabir, A., Yeasmin, D., Islam, M. A., Chowdhury, K. I. A., Anwar, K. S., Chughtai, A. A., & Seale, H. (2020). COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4), 1621-1629. <https://doi.org/10.4269/ajtmh.20-0812>
- Jacobs, L. G. (2022). Freedom of speech and regulation of fake news. *The American Journal of Comparative Law*, 70(Suppl_1), i278-i311. <https://doi.org/10.1093/ajcl/avac010>
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102-138). Guilford Press.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3-28. <https://doi.org/10.1037/0033-2909.114.1.3>
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339-375. <https://doi.org/10.1037/0033-2909.129.3.339>
- Judd, C. M., & Kulik, J. A. (1980). Schematic effects of social attitudes on information processing and recall. *Journal of Personality and Social Psychology*, 38(4), 569–578. <http://dx.doi.org/10.1037/0022-3514.38.4.569>
- Kensinger, E. A., & Schacter, D. L. (2008). Memory and emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 601-617). The Guilford Press.
- Koller, J. E., Villinger, K., Lages, N. C., Brünecke, I., Debbeler, J. M., Engel, K. D., Griebel, S., Homann, P. C., Kaufmann, R., Koppe, K. M., Oppenheimer, H., Radtke, V. C., Rogula, S., Stähler, J., Renner, B., & Schupp, H. T. (2021). Stigmatization of Chinese and

- Asian-looking people during the COVID-19 pandemic in Germany. *BMC Public Health*, 21, Article 1296. <https://doi.org/10.1186/s12889-021-11270-1>
- Kovach, B., & Rosenstiel, T. (2014). *The elements of journalism: What newspeople should know and the public should expect* (3rd ed.). Crown Publishing Group.
- Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J. B., Barzilai, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., ... Wineburg, S. (2022). *Toolbox of Interventions Against Online Misinformation*. <https://doi.org/10.31234/osf.io/x8ejt>
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790. <https://doi.org/10.1073/pnas.1320040111>
- Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, 64(4), 241-256. <https://psycnet.apa.org/doi/10.1037/a0015309>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International affective picture system (IAPS): Technical manual and affective ratings*. University of Florida. <https://acordo.net/acordo/wp-content/uploads/2020/08/instructions.pdf>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- Lee, S. K., Sun, J., Jang, S., & Connelly, S. (2022). Misinformation of COVID-19 vaccines and vaccine hesitancy. *Scientific reports*, 12(1), 13681. <https://doi.org/10.1038/s41598-022-17430-6>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., & Newman, E. J. (2020). *The debunking handbook 2020*. Disponível em <https://www.climatechangecommunication.org/debunking-handbook-2020/>
- Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, 32(2), 348-384. <https://doi.org/10.1080/10463283.2021.1876983>
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585-589. [https://doi.org/10.1016/S0022-5371\(74\)80011-3](https://doi.org/10.1016/S0022-5371(74)80011-3)
- Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, 5(1), 86-88. <https://doi.org/10.3758/BF03336715>
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4(1), 19-31. <https://doi.org/10.1037/0278-7393.4.1.19>
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4), 361-366. <https://doi.org/10.1101/lm.94705>

- McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509-516. <https://doi.org/10.1037/0003-066X.52.5.509>
- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60(2), 175-215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, 24(2), 184-197. <https://doi.org/10.2307/2786067>
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 851-861. <https://doi.org/10.1037/0278-7393.19.4.851>
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413-439. <https://doi.org/10.1111/j.1460-2466.2010.01488.x>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210-220. <https://doi.org/10.1016/j.pragma.2013.07.012>
- Mitchell, S. E., Sadikova, E., Jack, B. W., & Paasche-Orlow, M. K. (2012). Health literacy and 30-day postdischarge hospital utilization. *Journal of Health Communication*, 17(3), 325-338. <https://doi.org/10.1080/10810730.2012.715233>
- Moravec, P. L., Minas, R. K., & Dennis, A. R. (2018). *Fake news on social media: People believe what they want to believe when it makes no sense at all* (Kelley School of Business Research Paper No. 18-87). SSRN. <https://doi.org/10.2139/ssrn.3269541>
- Möttus, R., & Rozgonjuk, D. (2021). Development is in the details: Age differences in the Big Five domains, facets, and nuances. *Journal of Personality and Social Psychology*, 120(4), 1035-1048. <https://doi.org/10.1037/pspp0000276>
- Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, 67(4), 586-609. <https://doi.org/10.1111/jcom.12312>
- Neyazi, T. A., Kalogeropoulos, A., & Nielsen, R. K. (2021). Misinformation concerns and online news participation among internet users in India. *Social Media+ Society*, 7(2), 20563051211009013. <https://doi.org/10.1177/20563051211009013>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, 15(3), 259-267. <https://doi.org/10.1093/heapro/15.3.259>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330. <https://doi.org/10.1007/s11109-010-9112-2>
- Osmundsen, M., Bor, A., Vahlstrup, P., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review*, 115(3), 999-1015. <https://doi.org/10.1017/S0003055421000290>
- Papadogiannakis, E., Papadopoulos, P., P. Markatos, E., & Kourtellis, N. (2023). Who funds misinformation? A systematic analysis of the ad-related profit routines of fake news sites. *Proceedings of the ACM Web Conference 2023*, 2765-2776. <https://doi.org/10.1145/3543507.3583443>

- Pennycook, G., Bear, A., Collins, E., & Rand, D. G. (2020). The Implied Truth Effect: Attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. *Management Science*, 66(11), 4944-4957. <https://doi.org/10.1287/mnsc.2019.3478>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590-595. <https://doi.org/10.1038/s41586-021-03344-2>
- Piksa, M., Noworyta, K., Piasecki, J., Gwiazdzinski, P., Gundersen, A. B., Kunst, J., & Rygula, R. (2022). Cognitive processes and personality traits underlying four phenotypes of susceptibility to (mis) information. *Frontiers in Psychiatry*, 13, 912397. <https://doi.org/10.3389/fpsy.2022.912397>
- Lin, H., Pennycook, G., & Rand, D. G. (2023). Thinking more or thinking differently? Using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing. *Cognition*, 230, 105312. <https://doi.org/10.1016/j.cognition.2022.105312>
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, 48, 609-647. <https://doi.org/10.1146/annurev.psych.48.1.609>
- Polage, D. C. (2012). Making up history: False memories of fake news stories. *Europe's Journal of Psychology*, 8(2), 245-250. <https://doi.org/10.5964/ejop.v8i2.456>
- Rocha, Y. M., Moura, G. A., Desidério, G. A., Oliveira, C. H., Lourenço, F. D., & Nicolette, L. D. F. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review. *Journal of Public Health*, 31(7), 1007-1016. <https://doi.org/10.1007/s10389-021-01658-z>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1). <https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., & van der Linden, S. (2022). How to combat health misinformation: A psychological approach. *American Journal of Health Promotion*, 36(3), 569-575. <https://doi.org/10.1177/08901171211070958>
- Sabbagh, C., Boyland, E., Hankey, C., & Parrett, A. (2020). Analysing credibility of UK social media influencers' weight-management blogs: a pilot study. *International Journal of Environmental Research and Public Health*, 17(23), 9022. <https://doi.org/10.3390/ijerph17239022>
- Santos, M. M. M. T. D. (2023). Correção da Desinformação em Contextos de Saúde [Dissertação de mestrado, Iscte – Instituto Universitário de Lisboa]. Repositório do ISCTE-IUL. <http://hdl.handle.net/10071/29309>
- Saraiva, M., & Garrido, M. V. (2024). Misinformation in social interaction: examining the role of discussion. *Memory*, 32(2), 156-165. <https://doi.org/10.1080/09658211.2023.2300671>
- Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., Tate, C., & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology*, 40(4), 274. <https://doi.org/10.31234/osf.io/t7bw2>
- Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16), 7662-7669. <https://doi.org/10.1073/pnas.1805871115>
- Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, 2(1), 85-95. <https://doi.org/10.1353/bsp.2016.0009>
- Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PloS One*, 13(4), e0196087. <https://doi.org/10.1371/journal.pone.0196087>

- Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M., & Liu, H. (2020). Combating disinformation in a social media age. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(6), e1385. <https://doi.org/10.1002/widm.1385>
- Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review*, 12(3), 248-279. <https://doi.org/10.1177/1088868308319226>
- Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth. *Consciousness and Cognition*, 51, 53–67. <https://doi.org/10.1016/j.concog.2017.02.016>
- Singer, J. B. (2015). Out of bounds: Professional norms as boundary markers. In M. Carlson & S. C. Lewis (Eds.), *Boundaries of journalism: Professionalism, practices and participation* (pp. 21-37). Routledge. <http://dx.doi.org/10.4324/9781315727684>
- Southwell, B. G., Niederdeppe, J., Cappella, J. N., Gaysynsky, A., Kelley, D. E., Oh, A., ... & Chou, W. Y. S. (2019). Misinformation as a misunderstood challenge to public health. *American journal of preventive medicine*, 57(2), 282-285. <https://doi.org/10.1016/j.amepre.2019.03.009>
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133(3), 572-585. <https://doi.org/10.1016/j.cognition.2014.08.006>
- Tay, L. Q., Tan, J. J. X., & Ng, Y. Q. (2021). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology*, 113(3), 591-607. <https://doi.org/10.1111/bjop.12551>
- Unkelbach, C., & Rom, S. C. (2017). A referential theory of the illusory truth effect. *Journal of Experimental Psychology: General*, 146(4), 464-484. <https://doi.org/10.1016/j.cognition.2016.12.016>
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008. <https://doi.org/10.1002/gch2.201600008>
- van Prooijen, J. W., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, 48(7), 897-908. <https://doi.org/10.1002/ejsp.2530>
- Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The illusory truth effect leads to the spread of misinformation. *Cognition*, 236, 105421. <https://doi.org/10.1016/j.cognition.2023.105421>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, 39(5), 621-645. <https://doi.org/10.1177/1075547017731776>
- Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., & Walker-Smith, J. A. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103), 637-641. [https://doi.org/10.1016/S0140-6736\(97\)11096-0](https://doi.org/10.1016/S0140-6736(97)11096-0)
- West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences*, 118(15). <https://doi.org/10.1073/pnas.1912444117>
- Wright, D. B., & Loftus, E. F. (2001). How misinformation alters memories. *Journal of Experimental Child Psychology*, 71(2), 155-164. <https://doi.org/10.1006/jecp.1998.2467>

- Wu, C., Wu, F., Qi, T., & Huang, Y. (2020, July). User Modeling with Click Preference and Reading Satisfaction for News Recommendation. In *International Joint Conference on Artificial Intelligence* (pp. 3023-3029). <https://doi.org/10.24963/ijcai.2020/418>
- Yılmazdoğan, O. C., Şen Doğan, R., & Altıntaş, E. (2021). The impact of the source credibility of Instagram influencers on travel intention: The mediating role of parasocial interaction. *Journal of Travel & Tourism Marketing*, 38(8), 743–758. <https://doi.org/10.1177/1356766721995973>
- Zaragoza, M. S., & Lane, S. M. (1994). Source misattribution and the suggestibility of eyewitness testimony. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 934-945. <https://psycnet.apa.org/doi/10.1037/0278-7393.20.4.934>
- Zimmerman, T., Shiroma, K., Fleischmann, K. R., Xie, B., Jia, C., Verma, N., & Lee, M. K. (2023). Misinformation and COVID-19 vaccine hesitancy. *Vaccine*, 41(1), 136-144. <https://doi.org/10.1016/j.vaccine.2022.11.014>

Anexos

Materiais do Estudo Principal

Figura A1

Fonte credível masculina – Médico Instagram



Figura A2

Fonte credível masculina – Médico LinkedIn



Figura A3

Fonte credível feminina – Investigadora Instagram



Figura A4

Fonte credível feminina – Investigadora LinkedIn

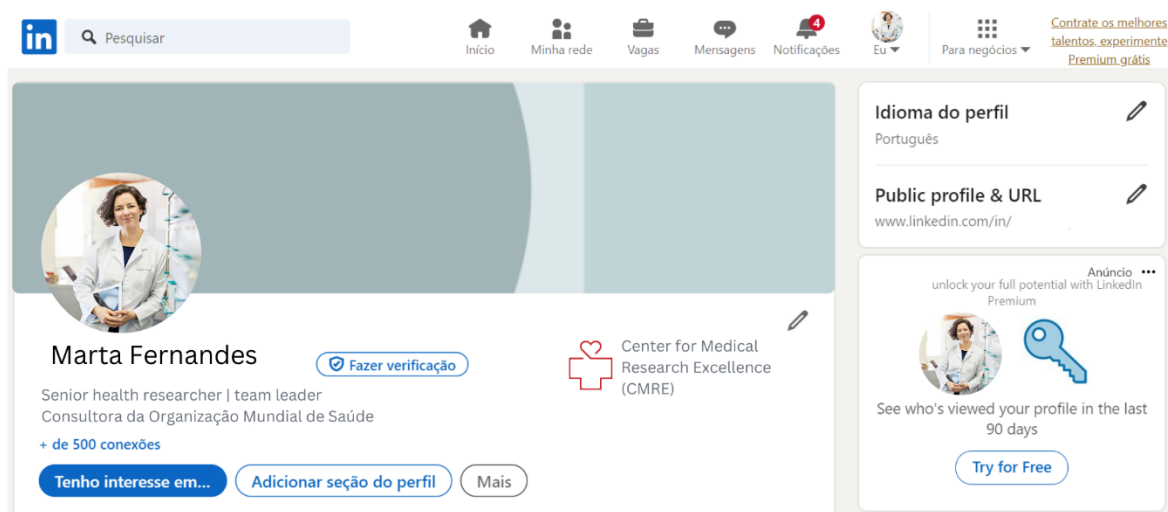


Figura A5

Fonte não credível masculina – Terapeuta holístico Instagram



Figura A6

Fonte não credível masculina – Terapeuta holístico LinkedIn

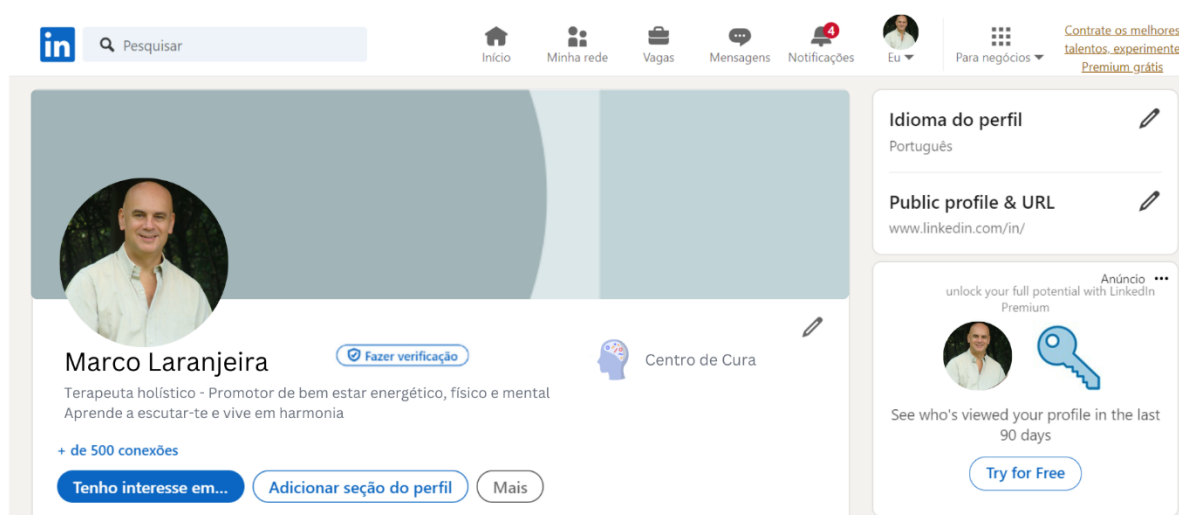


Figura A7

Fonte não credível feminina – Criadora de conteúdos Instagram



Figura A8

Fonte não credível feminina – Criadora de conteúdos LinkedIn

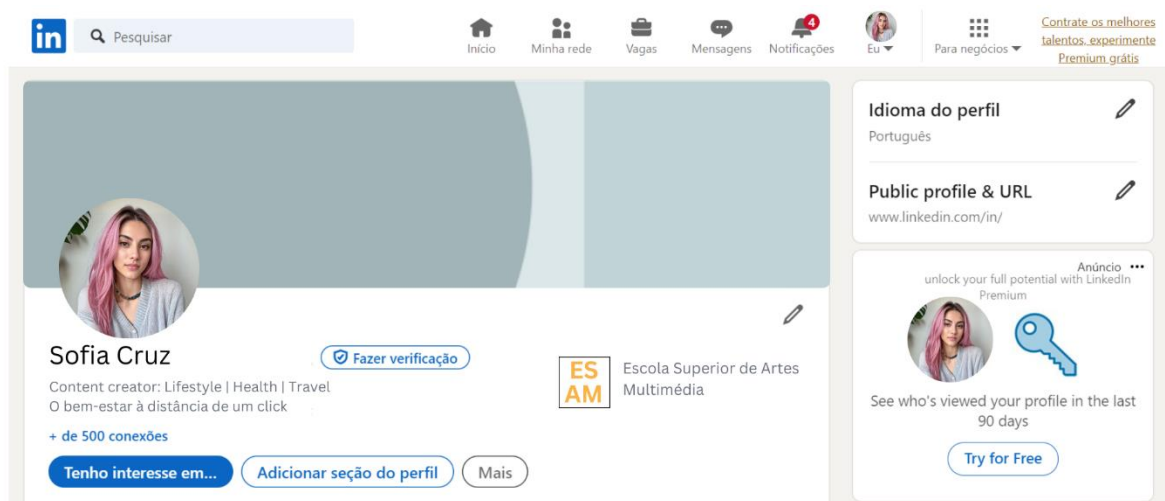


Figura A9

Fonte neutra Instagram

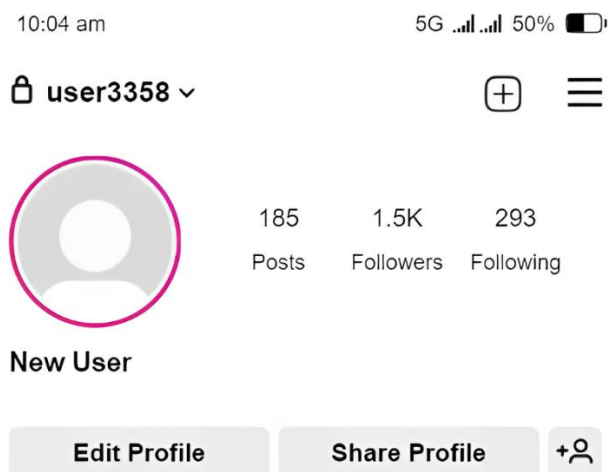


Figura A10

Fonte neutra LinkedIn

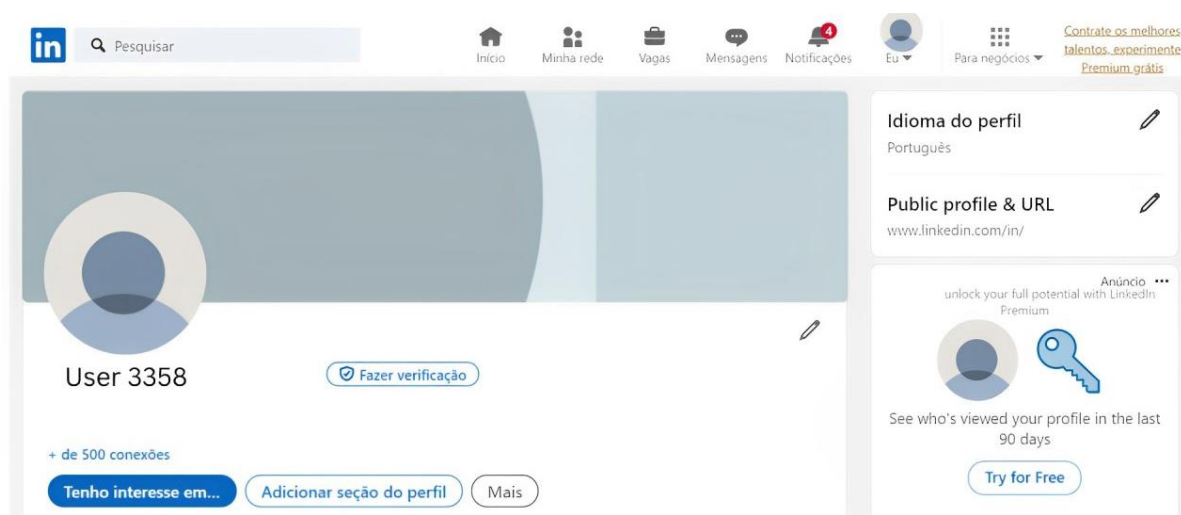


Tabela A1*Avaliação de veracidade das frases utilizadas no estudo (Santos, 2023)*

Frase	M
O açúcar mascavado é mais saudável que o branco	4.48
Comer cenouras melhora a visão	3.35
As causas do acne estão relacionadas com o consumo de chocolates	3.85
O leite é a maior fonte de cálcio	3.64
As dietas Detox permitem eliminar toxinas nocivas presentes no organismo	4.31
O vinho tinto previne as doenças cardíacas	3.57
A celulite pode ser eliminada com tratamento	4.32
Cortar o cabelo muito curto ou rapar ajuda a torná-lo mais forte	3.69
Uma dieta rica em vitamina D pode ajudar a recuperar dentes danificados com cáries	3.55
A restrição calórica de 30% aumenta a expectativa de vida nos mamíferos	3.91
O uso de células estaminais e de DNA aumenta o efeito dos cremes faciais	3.46
Pintar o cabelo durante a gravidez faz mal ao bebé	3.21
Existe uma relação entre o uso de telemóveis e o aparecimento de tumores no cérebro	3.54
Os campos magnéticos são prejudiciais à nossa saúde	4.39
Tomar vitaminas estimula o apetite sexual	3.35
Estalar os dedos pode provocar artrites e deformações	4.01
Os erros cometidos pelos médicos são a 3 ^a causa de morte	3.36
Conflitos e/ou traumas aumentam a probabilidade de se ter cancro	3.79
Quem toma a vacina da gripe pode contrair o vírus pela vacinação	3.54
Os medicamentos químicos provocam deficiências nutricionais	3.95
A pílula pode ser um fator promotor de cancro da mama, do útero e dos ovários	4.44
O Oscilloccinum é um “medicamento” homeopático comprovadamente eficaz na cura de gripes e constipações	3.87
Os antibióticos curam as constipações	3.43
De acordo com a Amnistia Internacional, a indústria farmacêutica tem preocupações maioritariamente humanitárias	3.62
As terapias não convencionais não são perigosas, na pior das hipóteses não fazem nenhum efeito	3.94
O corpo é percorrido por energias que podem ser manipuladas por agulhas, ímanes e pedras para reestabelecer o equilíbrio	3.98
A Medicina Quântica estimula as capacidades autocurativas do organismo	3.84
Os princípios da homeopatia são cientificamente fundamentados	3.70
A Energia Bioquântica revitaliza as células, evita e ajuda a tratar doenças	3.84
O Feng Shui usa as correntes de energia, e a disposição das coisas em casa, criando uma harmonia terapêutica eficaz	4.20

Nota: no estudo de Santos (2023) a veracidade das frases foi medida numa escala de 7 pontos (1= completamente falsa e 7 = completamente verdadeira)

Tabela A2*Lista de frases utilizadas e respectivas versões*

Versão 1		
Categoria	Repetidas	V/F
Alimentação	Comer cenouras melhora a visão	F
Alimentação	As causas do acne estão relacionadas com o consumo de chocolates	F
Estética	Uma dieta rica em vitamina D pode ajudar a recuperar dentes danificados com cáries	F
Estética	A restrição calórica de 30% aumenta a expectativa de vida nos mamíferos	F
Saúde	Os campos magnéticos são prejudiciais à nossa saúde	F
Saúde	Os erros cometidos pelos médicos são a 3ª causa de morte	F
Medicamentos	Os medicamentos químicos provocam deficiências nutricionais	F
Medicamentos	Os antibióticos curam as constipações	F
Terapias	As terapias não convencionais não são perigosas, na pior das hipóteses não fazem nenhum efeito	F
Terapias	O corpo é percorrido por energias que podem ser manipuladas por agulhas, ímanes e pedras para reestabelecer o equilíbrio	F
Contradições		
Alimentação	O açúcar mascavado não é mais saudável que o branco	V
Alimentação	O vinho tinto não previne doenças cardíacas	V
Estética	Não existe tratamento para eliminar a celulite	V
Estética	O uso de células estaminais e DNA não aumenta o efeito dos cremes faciais	V
Saúde	Não existe relação entre o uso de telemóveis e o desenvolvimento de tumores no cérebro	V
Saúde	As vitaminas não estimulam o apetite sexual	V
Medicamentos	As vacinas não causam as doenças para as quais oferecem proteção	V
Medicamentos	A pílula pode ser um fator protetor de cancro da mama, do útero e dos ovários	V
Terapias	A Medicina Quântica não estimula as capacidades autocurativas do organismo	V
Terapias	A Energia Bioquântica é uma pseudociência que não evita nem trata doenças	V
Novas		
Alimentação	O leite não provoca obesidade	V
Alimentação	Um treino intensivo não elimina calorias consumidas em excesso	V
Estética	Comer placenta pode ser perigoso para a saúde da mãe e para o bebé	V
Estética	A rifampicina é o único antibiótico que corta o efeito da pílula	V
Saúde	É impossível curar doenças pelo toque das mãos e canalização de energia (Reiki)	V
Saúde	O consumo de açúcar provoca hiperatividade nas crianças	F
Medicamentos	Os cremes e produtos de higiene têm químicos altamente perigosos para a saúde	F
Medicamentos	O cancro também é uma doença oncop psicológica e não só física	F
Terapias	Se tomar a pílula durante muitos anos pode ficar infértil	F
Terapias	Os suplementos alimentares são produtos naturais e por isso bons para a saúde	F

Versão 2		
Categoria	Repetidas	V/F
Alimentação	O açúcar mascavado é mais saudável que o branco	F
Alimentação	O vinho tinto previne as doenças cardíacas	F
Estética	A celulite pode ser eliminada com tratamento	F
Estética	O uso de células estaminais e de DNA aumenta o efeito dos cremes faciais	F
Saúde	Existe uma relação entre o uso de telemóveis e o aparecimento de tumores no cérebro	F
Saúde	Tomar vitaminas estimula o apetite sexual	F
Medicamentos	Quem toma a vacina da gripe pode contrair o vírus pela vacinação	F
Medicamentos	A pílula pode ser um fator promotor de cancro da mama, do útero e dos ovários	F
Terapias	A Medicina Quântica estimula as capacidades autocurativas do organismo	F
Terapias	A Energia Bioquântica revitaliza as células, evita e ajuda a tratar doenças	F
Contradições		
Alimentação	O leite não é a maior fonte de cálcio	V
Alimentação	O nosso corpo tem mecanismos de desintoxicação, dispensando dietas Detox	V
Estética	Pintar o cabelo durante a gravidez não faz mal ao bebé	V
Estética	Rapar ou cortar o cabelo curto não influencia a sua força ou crescimento	V
Saúde	Estalar os dedos não provoca artrites nem deformações	V
Saúde	Conflitos e/ou traumas não aumentam a probabilidade de ter cancro	V
Medicamentos	O Oscillococcinum não é um medicamento, é um produto homeopático inútil a tratar doenças	V
Medicamentos	De acordo com a Amnistia Internacional, a indústria farmacêutica só tem preocupações com os lucros económicos	V
Terapias	Os princípios da homeopatia não possuem fundamento científico	V
Terapias	O Feng Shui e a sua filosofia sobre as correntes de energia, e a disposição das coisas em casa é uma fraude	V
Novas		
Alimentação	Não existe uma quantidade fixa de água a beber por dia	V
Alimentação	Os cremes e produtos de higiene são altamente regulados e não têm químicos perigosos	V
Estética	O Calcitrin funciona tão bem como um placebo	V
Estética	Os medicamentos genéricos são mais baratos porque não pagam patente	V
Saúde	A fitoterapia usa ervas, mas não é segura para tratar quaisquer doenças	V
Saúde	Comer muito próximo das horas de dormir engorda	F
Medicamentos	A dieta Paleo é saudável promovendo uma alimentação que previne a mortalidade	F
Medicamentos	Ao longo do dia apenas usamos 10% do nosso cérebro	F
Terapias	Todos os antibióticos cortam o efeito da pílula anticoncecional	F
Terapias	A quiropraxia é um tratamento alternativo não perigoso	F

Versão 3		
Categoria	Repetidas	V/F
Alimentação	O leite é a maior fonte de cálcio	F
Alimentação	As dietas Detox permitem eliminar toxinas nocivas presentes no organismo	F
Estética	Pintar o cabelo durante a gravidez faz mal ao bebê	F
Estética	Cortar o cabelo muito curto ou rapar ajuda a torná-lo mais forte	F
Saúde	Estalar os dedos pode provocar artrites e deformações	F
Saúde	Conflitos e/ou traumas aumentam a probabilidade de se ter cancro	F
Medicamentos	O Oscillococcinum é um “medicamento” homeopático comprovadamente eficaz na cura de gripes e constipações	F
Medicamentos	De acordo com a Amnistia Internacional, a indústria farmacêutica tem preocupações maioritariamente humanitárias	F
Terapias	Os princípios da homeopatia são cientificamente fundamentados	F
Terapias	O Feng Shui usa as correntes de energia, e a disposição das coisas em casa, criando uma harmonia terapêutica eficaz	F
Contradições		
Alimentação	Comer cenouras não melhora a visão	V
Alimentação	As causas do acne são várias, mas comer chocolate não é uma delas	V
Estética	Uma dieta rica em vitamina D pode ajudar a prevenir caries nos dentes das crianças	V
Estética	A restrição calórica de 30% não tem influência na expectativa de vida	V
Saúde	Os campos magnéticos não são prejudiciais à saúde	V
Saúde	Erros médicos são um problema, mas não a 3ª causa de morte	V
Medicamentos	Medicamentos químicos não provocam deficiências nutricionais	V
Medicamentos	Os antibióticos curam as infeções	V
Terapias	As terapias não convencionais são perigosas, e podem mesmo ter graves impactos na saúde	V
Terapias	A ciência moderna não recorre a energias para explicar o funcionamento ou equilíbrio do corpo	V
Novas		
Alimentação	Comer banana não engorda	V
Alimentação	O suor consiste na perda de água e não contribui para emagrecer	V
Estética	A Cannabis não induz alterações da sexualidade	V
Estética	A toma da pílula não tem qualquer impacto na fertilidade	V
Saúde	Os suplementos alimentares mesmo sendo naturais podem ser prejudiciais à saúde	V
Saúde	Os alimentos sem glúten são recomendados a todos e não só a pessoas com intolerância	F
Medicamentos	Usar desodorizante antitranspirante pode causar cancro na mama	F
Medicamentos	Continuar a dar reforços da vacina para a COVID-19 só é vantajoso para as farmacêuticas	F
Terapias	Os medicamentos usados para as doenças mentais causam dependência	F
Terapias	A fitoterapia usa ervas e por isso é segura para tratar diversas doenças	F