



THE CONTINUED INFLUENCE EFFECT OF MISINFORMATION IN THE CONTEXT OF FORENSIC PSYCHOLOGY

Mikołaj BUCZEL, Adam SIWIAK, Malwina SZPITALAK

Institute of Psychology, Jagiellonian University, Kraków, Poland

Abstract

The continued influence effect (CIE) is a phenomenon of the continuous influence of misinformation on the inferences of an individual, despite the misinformation being retracted. The aim of this article is to present this phenomenon, compare it to accounts of the post-event misinformation effect (PME) in the Polish literature and indicate the practical implications of research into it. The thirty-year history of research into the continued influence effect, which has investigated its mechanisms and determinants, as well as ways to reduce it, is reviewed. An attempt is also made to organise the nomenclature related to misinformation, focusing on the native meaning of this word. As eyewitness testimonies are prone to distortion and are often retracted or corrected during investigations, various judicial authorities can be exposed to the effects of CIE, such as improper decisions and unfair sentences. Attending to the research on CIE and its forensic context can reveal ways of minimising the effects of the phenomenon.

Keywords

Misinformation; Eyewitness testimony; Post-event misinformation effect; Continued influence effect; Forensic psychology.

Received 17 November 2020; accepted 7 December 2020

Misinformation – origins of the concept and definitions

The term “disinformation” is probably of Russian origin (*дезинформация; dezinformatsiya*) and in this sense refers to the deliberate dissemination of false information with the intention of influencing the attitudes of those whom the information reaches (Prokhorov, 1974–1983). English also has a much older term, dating to the 16th century (Online Etymology Dictionary, n.d.), *misinformation*, meaning false or inaccurate information which, unlike *disinformation*, is not deliberately disseminated (cf. “disinformation”, Online Etymology Dictionary, n.d.; “misinformation”, Merriam-Webster’s collegiate dictionary, n.d.). The Polish language lacks this distinction. This seemingly trivial problem may lead to an incomplete or incorrect picture of the phenomenon of misinformation in areas

other than those most often mentioned by the Polish literature. Here, misinformation is treated, for example, as a form of communication that misleads recipients and interferes with their decision-making processes (Basaj, 2018), as a component of propaganda (Rajczyk, 2016) or as misinformation in the absence of reliable information (Dubisz, 2003), which refers to the media, the Internet environment and social media (*fake-news*; Brzóska, Nowak, 2019), as well as to actors within the military, economic, scientific-technological and political communities (Wachowicz, 2019). These reviews consistently fail to discuss misinformation in the psychological sense, which may be, among other things, a consequence of the use of the expression *misinformation* rather than *disinformation* in the foreign literature. In the Polish psychological literature, on the other hand, we can find information about the misinformation effect, defined as inclusion

by the person making a memory report of information that is inconsistent with the course of the event and comes from sources other than the event itself (Polczyk, 2007). In the English-language literature, in addition to the aforementioned misinformation effect, researchers also deal with the so-called continued influence effect (CIE; Johnson, Seifert, 1994), which denotes the continued influence of items of information on accounts and conclusions about an event. Significantly, this information, even though it has been corrected and acknowledged as incorrect, continues to be reproduced and influences the judgments of those exposed to it. Although it is a popular topic, the Polish literature lacks experimental research in this area. This paper aims to present and compare the two phenomena mentioned above.

Post-event misinformation effect (PME)

The misinformation effect is usually considered in terms of the reconstructive nature of memory (Mitchell, Johnson, 2000). In the majority of cases, it has been studied in the context of the law courts (Loftus, 2005), where witnesses are exposed to the reception of information coming to them from various sources other than the event itself. If the information is false, witnesses' memory reports – although unintentionally and in good faith – may be distorted and have implications for litigation.

Characteristic of PME is its three-stage course, which is present both in non-laboratory conditions and mapped in research procedures (e.g. Loftus, Miller, Burns, 1978). In the first stage, the respondents are introduced to the original material (e.g. an extract from a film, a text read aloud or an enactment), and after some time, to the follow-up material (e.g. a description of an event or another person's account), which contains misinformation. In the final stage, the respondents answer questions about the original material. The differences obtained in experimental conditions similar to those described above between misinformed and not-misinformed groups usually reach a value of several tens of percent and are perfectly replicable (cf. Polczyk, 2007).

Misinformation effect mechanisms

While the fact of the existence of PME is generally not in dispute, its mechanisms remain more problematic. In the early days of research into this area, concepts of errors in the functioning of the memory sys-

tem were almost exclusively accepted (Loftus et al., 1978). The earliest of these claimed that the memory trace of the original information is, due to the presence of misinformation, damaged (Loftus, 1975) or even overwritten (Loftus et al., 1978). Later theories have focused on the coexistence of both memory traces, of which the one containing misinformation worsens access to the original (Bekerian, Bowers, 1983), on the more rapid disintegration of a representation of the form of an event while retaining its content, which can thus be generalised to misinformed details (Brainerd, Reyna, 2002), or on the co-occurrence of associations and errors in identifying their source (as a consequence of misinferring the source of activation of a particular representation; Ayers, Reder, 1998). However, it seems that the most widely accepted and contemporarily dominant theory is that of information-source monitoring error (Lindsay, Johnson, 1989a), which considers incorrect attribution of information source to information content as the cause of PME. However, this theory has not proven to be complete (Polczyk, 2007). There are a number of studies indicating that memory reports can also be inconsistent with the original material when the memory of the original material remains intact (e.g. Blank, 1998; Polak, Dukala, Szpitalak, Polczyk, 2016; Polczyk, 2007; 2017). Accordingly, some researchers suggest that detail memory and source memory should be considered as two separate manifestations of PME (Pansky, Tenenboim, Bar, 2011).

The first studies to question the purely memory-based mechanisms of the misinformation effect were published by McCloskey and Zaragoza (1985), but systematic exploration of non-memory mechanisms began with Blank's research (1998). By using a four-step procedure involving the addition of a second final test, preceded by informing the subjects of the existence of a discrepancy between the original and follow-up material, Blank found that, on average, approximately 40% of the subjects who were misinformed were aware of the discrepancy, which was mainly explained by the subjects' own memory uncertainty.

Continued influence effect (CIE)

With the widespread impact of misinformation on various aspects of individual and larger social group functioning, there is a need to effectively correct it. This task is difficult because disinformation, once assimilated, cannot be easily removed. The pioneers in the study of this phenomenon, referred to as the

continued influence effect, were Johnson and Seifert (1994), and Wilkes and Leatherbarrow (1988). The classic CIE research paradigm involves presenting participants with a scenario of a fictitious incident, followed by a so-called retraction of information about the cause of the incident, i.e., informing respondents that in fact the information was not true. For example, respondents were presented with fictitious reports of a fire on premises where paper materials were located. The first version stated that flammable materials were stored in the room where the incident occurred without sufficient safety precautions. Then this information was contradicted by stating that no such material was there. Despite the memory of the retraction of the information, the respondents in the final test continued to form conclusions that included information about flammable materials. This effect is well replicable, and in most studies a significant proportion of subjects use misinformation in their judgments (e.g. Ecker, Ang, 2019; Ecker, Hogan, Lewandowsky, 2017; Ecker, Lewandowsky, Swire, Chang, 2011).

Mechanisms of the continued influence effect

There are two dominant, though not mutually exclusive, cognitive explanations of CIE mechanisms in the literature. The authors of the first suggest that it is the result of an error during information retrieval (e.g. Ecker et al., 2010), and assume that misinformation and its reference are stored within different representations in memory and that they compete for activation during retrieval. Misinformation affects inference when it, not the correction, is selectively chosen. This concept combines the operation of both automatic and strategic retrieval and monitoring processes: when misinformation is activated automatically by cues, strategic processes are required to assess the validity of the retrieved information and/or recall any relevant corrections. Because strategic processes are cognitively demanding, they often fail, which results in a continued influence effect.

An alternative explanation is offered by the theory of mental models (Ecker et al., 2011), which is based on the way people process information about events. Mental models, built from extracted content, consist of elements linked together in a cause-and-effect chain. Withdrawing information from the resulting model creates a gap that makes it less complete. Misinformation, though withdrawn, can fill this gap, allowing the construction of a complete, though potentially incorrect, model, which is nonetheless preferable to an accurate but incomplete one (Ecker et al.,

2011; Walter, Tukachinsky, 2020). Some behavioural and neuropsychological evidence supports the theory of mental models: fMRI neuroimaging studies have shown that the activity of neurons associated with information integration is different according to whether there has been a retraction of prior information or no such retraction, which may suggest that a failure to integrate and update memory is behind CIE (Gordon, Brooks, Quadflieg, Ecker, Lewandowsky, 2017). Other studies show that integration is more likely to fail when an individual does not make full use of their cognitive resources, which may, for example, be due to cognitive load when encoding the retraction of information at the same time as attempting to integrate it (Ecker et al., 2011).

Hamby, Ecker and Brinberg (2020) find that the ineffectiveness of the retraction of misinformation has to do with the relationship between the key piece of information and the broader context (what the individual had in memory previously), and with the fact that retracted misinformation has a stronger effect on an individual's reasoning and beliefs if it can retrospectively explain the story that the individual previously had in memory. Participants in the experiment were presented with a story about a person who experienced complications after taking a certain drug. Some of the subjects were given the reason for these complications in the explanations and another group was left without explanations. Both groups were then given another piece of information that could potentially explain these complications, after which the information was withdrawn. During the post-test, it was found that those deprived of initial information about the reasons for the complication further succumbed to the misinformation by using it to make inferences. This misinformation helped fill a gap in their mental model. Respondents who were initially given an explanation about the reasons for the complication, with no gap to fill, were less likely to rely on the withdrawn misinformation.

Determinants of the continued influence effect

Although cognitive factors play a significant role in the operation of the continued influence effect, they cannot completely explain it. Motivational factors such as attitudes or world-view are also considered in the literature. It is worth noting that much of the research into CIE refers to how it affects an individual's reasoning and judgments, with less attention paid to the actual impact on people's behaviours and choices. In contrast, research findings (Swire, Berin-

sky, Lewandowsky, Ecker, 2017) show that correcting misinformation does not necessarily translate into behaviour change: feelings towards politicians who disseminated misinformation remained unchanged even when participants acknowledged that the statements they had made had been incorrect; and people still intended to vote for them. This point is made with reference to lies by Briony Swire-Thompson and colleagues (2020). Ecker and Ang (2019), studying supporters of right-leaning and left-leaning parties in Australia, found that partisan attitudes affect retraction processing, especially if the misinformation relates to general views pertaining to a particular political wing – retraction of misinformation is more effective if it does not require a change in political attitudes. Misinformation is especially difficult to retract if it aligns with conservative views. Adjustments inconsistent with world-view may even, under certain conditions, be counterproductive (backfire), i.e., paradoxically increase submission to misinformation (Lewandowsky, Ecker, Seifert, Schwarz, Cook, 2012; results to the contrary: Ecker et al., 2014).

The role of world-view, and of personal likes or biases, is not limited to succumbing to misinformation on political issues. In a study by Ecker et al. (2014) participants with high and low levels of racial prejudice were presented with an event involving a fictional robbery in which an Australian Aboriginal was a suspect in one version and the hero who thwarted the robbery in another. It appeared that racial views did not determine the effectiveness of retraction, although it was noticeable that those with a high degree of racial bias were more likely to mention the Aboriginal suspect in their statements than the Aboriginal hero. Similar results were obtained in a study in which a fictional plane crash was first attributed to a terrorist attack and then a correction was made to explain that the plane had had a damaged fuel tank. Individuals with high levels of islamophobia on the inference test were more likely to allude to terrorist themes than those with low levels, although retraction was equally effective in both groups. Interestingly, this study also found that the emotionality of a story does not affect the degree to which misinformation is digested, and that emotionally-charged information is no more resistant to correction than less emotionally-charged information (Ecker, Lewandowsky, Apai, 2011).

Suspicion, either of the source of the information or of the intentions of the participants in an event, is a factor that potentially reduces CIE. Research into the 2003 Iraq War conducted with three groups of participants, two from coalition countries (the US and Australia) and the third from Germany, which opposed the

war, showed that participants from the US were the least responsive to correction of misinformation related to the event. Germans, being the most sceptical about the official reason for the war (the dismantling of weapons of mass destruction), were less likely to succumb to misinformation (Lewandowsky et al., 2005). Nevertheless, in many situations it may not be possible to arouse suspicion because it is associated with scepticism, which is a stable personality trait (Lewandowsky, Stritzke, Oberauer, Morales, 2005; 2009).

The repetition of misinformation in retractions of misinformation is also analysed (Ecker et al., 2017; 2011). There is an ongoing debate among researchers about whether misinformation should be repeated in corrections. A fairly common view is that repeating misinformation in retractions can cause the so-called familiarity effect, whereby something that sounds familiar is more likely to be recognised as true (Swire, Ecker, Lewandowsky, 2017). In the context of CIE, misinformation repeated in corrections could come to mind more easily in the future, e.g. explaining that vaccines do not cause autism includes the association of “vaccine” and “autism”, making the false statement about the negative effects of vaccines sound familiar (Lewandowsky et al., 2012; Swire et al., 2017). However, the results of some studies show the opposite effects (Ecker et al., 2017), suggesting that repeating misinformation in corrections can be considered an effective and safe strategy. At the same time, the researchers expressed some caveats: firstly, the interval between the presentation of misinformation and its correction was shorter than would be the case in many real-world situations (time, in turn, can affect the effectiveness of corrections; Walter, Tukachinsky, 2020) and secondly, corrections containing alternative causes are often not available in the real world. It is possible that repetition of misinformation can only have a negative effect in the absence of such alternatives (Ecker et al., 2017). However Swire et al. (2017) found no evidence for this. Ecker, O’Reilly, Reid and Chang (2020), on the other hand, suggest that when subjects become familiar with previously unknown misinformation, there may nevertheless be a familiarity effect and an associated danger of spreading it through the population, but the results of another study (Ecker et al., 2020) show that even when the misinformation was not previously known, its repetition in the correction did not increase the continued influence effect.

Immunisation techniques

A very important issue in the area of CIE research from a practical perspective is methods of immunisation against the influence of misinformation. The consequences of succumbing to false information spread in public spaces can be seen, for example, in the effects of fake-news (Lazer et al., 2018) or the belief that vaccines cause autism (Poland, Spier, 2010). These examples illustrate well the problem that simply retracting or correcting false information does not stop its influence on an individual's views. So far, three factors have been identified to reduce the impact of misinformation in the CIE paradigm: warning, repeated retraction, and presentation of alternatives.

Presenting an alternative – one of the above-mentioned techniques (based on the aforementioned theory of mental models) is to present an alternative cause for an event, which replaces previous misinformation (Ecker et al., 2010; Johnson, Seifert, 1994), and to highlight the contradiction between legitimate and illegitimate information, which significantly facilitates integration/updating processes. It is believed that people prefer to retain key information in their memories, that is, information that relates to the causes of events and the people involved, and that providing an explanation other than misinformation aids the process of correct inference while filling a gap in the mental model without compromising its reliability. Ecker et al. (2011) also manipulated the strength of the correction of misinformation by repeating corrective information several times, which they believe can lead to stronger mental model updating immediately after encoding of a correction. Although rebuttal and repeated correction techniques are more effective than correction alone, CIE is never completely removed.

Warning of the possibility of misinformation has been studied quite extensively in the PME paradigm (e.g., Blank, Launay, 2014; Chambers, Zaragoza, 2001; Greene, Flynn, Loftus, 1982; Szpitalak, 2015; Wright, 1993). It is a technique that is simple to induce, ecologically accurate and effective regardless of location (cf. Szpitalak, 2015). However, the study conducted by Ecker et al. (2010) was probably the first to use a warning about misinformation in the continued influence effect paradigm. In their experiments, they presented subjects with a fictional account of an accident involving a minibus that, according to the first report, was travelling with elderly nursing home residents. In subsequent steps, depending on the group, (1) the information was retracted or (2) replaced with other information (stating that the passengers were players on the student hockey team), (3) the informa-

tion was retracted and subjects were given a pre-written warning that sometimes first reports of events may be imprecisely authenticated and untrue, and (4) subjects were informed of the existence of CIE with specific examples of how it works. The general warning in condition (3) was effective only for those subjects who remembered the retraction of the misinformation, whereas the manipulation in condition (4) reduced dependence on the misinformation to a degree similar to giving a plausible alternative. The next experiment combined the two most effective manipulations, that is, giving a plausible alternative with a simultaneous warning about CIE. A reduction in succumbing to misinformation was achieved to a greater extent than when using any of the techniques on their own, nevertheless the effect was not completely eliminated (Ecker et al., 2010).

Practical implications and further research directions

Despite more than forty years of psychological research on the discussed misinformation effects, there have been very few attempts to integrate them (cf. Ecker et al., 2015). This seems rather surprising, especially given the similarities between experiments traditionally included in the PME trend and the CIE phenomenon (e.g. Lindsay, Johnson, 1989b).

One of the benefits of studying PME and CIE together would be the ability to gain a deeper understanding of both phenomena and create more effective immunisation techniques, especially since those developed to date appear to be suitable for use in non-laboratory settings to varying degrees. In the case of PME, many of these are rather rudimentary (although many have shown high efficacy and applicability in practice, e.g. Greene et al., 1983; Memon, Meissner, Fraser, 2010; Szpitalak, 2012; Szpitalak, Polczyk, 2012; 2019), while the repertoire of CIE immunisation methods seems very poor and not very well studied. Theoretical speculation regarding the application of PME immunisation techniques to CIE seems promising: as noted by Szpitalak and Polczyk (2017), a very important step towards improving immunisation methods would be to test them when the individual is aware of inconsistencies between the information reaching them. So far, this has been done for warning (Polczyk, 2007) and enhanced self-affirmation (Szpitalak, 2012), indicating that the methods are more effective precisely when the individual has memories of both the original information and the misinformation, which appears also to

be the case for giving detailed information about the continued influence effect (Ecker et al., 2010).

In contrast, existing knowledge and new research on CIE could be of considerable importance to applied forensic psychology. Information that is often disclosed in the media before trials threatens the impartiality of the judiciary, which is why it is so important to avoid disclosing certain categories of information before they begin. When information of this kind introduces doubt about defendants' innocence or portrays defendants in a negative light, it is possible that they will be found guilty and convicted more often. For example, by presenting subjects with false, guilt-enhancing information about a person accused of a crime, Gilbert, Tafarodi and Malone (1993) showed that despite experimental subjects reading true information contradicting the defendant's guilt, they judged that the defendant deserved a more severe prison sentence. Black (2018), on the other hand, suggests the relevance of CIE immunisation to witness retraction and judicial bias: as mentioned when discussing PME, eyewitness testimony can be distorted and then, despite being retracted, continue to influence judicial decisions, especially where the judiciary is unprofessional, biased or prejudiced. The author also suggests refraining from reporting the retraction of testimony until alternative explanations of the case are available. It is also possible to link CIE to the problem of false self-incrimination: when acquainting themselves with the testimony of people who were coerced into confessing during interrogation, judges made convictions just as often as they did against people who confessed without coercion. This happened despite the judges' declarations that the fact that the testimony was coerced would not affect their decisions (Kassin, Sukel, 1997). On the other hand, Jones, Crozier and Strange (2017), who combined the PME and CIE paradigms, presented results showing that even with direct evidence, people are still capable of being guided by inferences consistent with misinformation. Respondents read a description of an altercation between a citizen and a police officer and how it ended and read a misinformation-laden police report about the incident, which blamed the citizen and excused the police officer. They then watched a video shot from the body-worn camera that the same police officer had been wearing. It appeared that despite viewing a direct recording of the event, respondents generated inferences consistent with the police misinformation and hence blamed the citizen.

What is more, the instruction received by juries as to which evidence and testimony is ultimately admitted and which is rejected offers insufficient protec-

tion against CIE and usually proves ineffective (Fein, McCloskey, Tomlinson, 1997). Before agreeing on a verdict, the judge tells the jurors which evidence appearing in the investigation was not admitted at trial which, in the continued influence effect, would correspond to the retraction of false information (Fein, McCloskey, Tomlinson, 1997). Juries often ignore such instructions and make their decisions based on the retracted evidence, which is a manifestation of the continued influence effect. Steblay, Hosch, Culhane, and McWethy (2006) report from their study that if in their instructions judges explain to juries why a particular piece of evidence was retracted (the equivalent of reliable alternative information), they diligently follow the instruction (the equivalent of reducing the continued influence effect).

One of the remedies for CIE that courts sometimes use is to postpone trials. If pre-trial media coverage of a case presents information suggesting a defendant's guilt, the defence may move for postponement in the hope that the media hype will fall silent. Research findings suggest that such a delay may have positive results (Kramer, Kerr, Carroll, 1990). On the other hand, Ecker et al. (2010) find that warnings appear to be a promising and economical way to prevent judges from relying on already outdated information in making judgments, and suggest that specific warnings should be routinely given to judges in advance to arouse their suspicion about the information coming to them. Similar conclusions can be drawn from older studies. Fein et al. (1997), for example, showed that the impact of retracted false evidence and accusations on judges can be reduced if, prior to their presentation, subjects were made suspicious of the reasons why the information was being made available.

In the case of witness testimony, the CIE schema seems to complement that of the PME: when a witness to an event encounters misinformation, which is then retracted, they may still rely on it in their testimony. This is strikingly reminiscent of the negated misinformation procedure used in the PME research strand (Lindsay, 1990), in which the subject is informed that one detail in the original material was described incorrectly and that the follow-up material does not contain the correct answers to the final test questions. Existing knowledge of CIE may thus prove invaluable to witness testimony: the use of alert, rebuttal, or repeated retraction procedures, as well as of potential new techniques, could reduce the chance of succumbing to misinformation when someone does not directly witness an event or, especially, when someone does not remember an event but nevertheless testifies based

on information inferred from news reports, rumours or conversations with others.

References

1. Ayers, M. S., Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, 5(1), 1–21.
2. Basaj, K. (2018). Dezinformacja czy sztuka manipulacji. *Biuletyn Biura Analiz i Reagowania Rządowego Centrum Bezpieczeństwa*, 25, 14–17.
3. Bekerian, D. A., Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 139–145.
4. Black, S. C. (2018). The continued influence of retracted witness testimony at jury trials. *Asia Pacific Journal of Advanced Business and Social Studies*, 4(1), 283–290.
5. Blank, H. (1998). Memory states and memory tasks: An integrative framework for eyewitness memory and suggestibility. *Memory*, 6(5), 481–529.
6. Blank, H., Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, 3(2), 77–88.
7. Brainerd, C. J., Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11(5), 164–169.
8. Brzóska, P., Nowak, B. (2019). Fałsz w erze mediów społecznościowych, czyli różnice indywidualne, bańki informacyjne i podatność na fake newsy. *Badania i Rozwój Młodych Naukowców w Polsce. Nauki humanistyczne i społeczne. Część V – Społeczeństwo i rozwój*, 21–27.
9. Chambers, K. L., Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory and Cognition*, 29, 1120–1129.
10. Dalton, A. L., Daneman, M. (2006). Social suggestibility to central and peripheral misinformation. *Memory*, 14(4), 486–501.
11. Disinformation. (In) *Merriam-Webster's collegiate dictionary*. Retrieved September 7, 2020 from: <https://www.merriam-webster.com/dictionary/disinformation>.
12. Dubisz, S. (Ed.). (2003). *Uniwersalny słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
13. Ecker, U. K. H. (2017). Why rebuttals may not work: the psychology of misinformation. *Media Asia*, 44(2), 79–87.
14. Ecker, U. K. H., Ang, L. C. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology*, 40(2), 241–260.
15. Ecker, U. K. H., Hogan, J. L., Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185–192.
16. Ecker, U. K. H., Lewandowsky, S., Apai, J. (2011). Terrorists brought down the plane! – No, actually it was a technical fault: Processing corrections of emotive information. *The Quarterly Journal of Experimental Psychology*, 64(2), 283–310.
17. 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), 2–25.
18. Ecker, U. K. H., Lewandowsky, S., Fenton, O., Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, 42(2), 292–304.
19. Ecker, U. K. H., Lewandowsky, S., Swire, B., Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin and Review*, 18(3), 570–578.
20. Ecker, U. K. H., Lewandowsky, S., Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory and Cognition* 38(8), 1087–1100.
21. Ecker, U. K. H., O'Reilly, Z., Reid, J. S., Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, 111(1), 36–54.
22. Fein, S., McCloskey, A. L., Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin*, 23(11), 1215–1226.
23. Gilbert, D. T., Tafarodi, R. W., Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65(2), 221–233.
24. Gordon, A., Brooks, J. C. W., Quadflieg, S., Ecker, U. K. H., Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing. *Neuropsychologia*, 106, 216–224.
25. Greene, E., Flynn, M. B., Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Learning and Verbal Behavior*, 21, 207–219.
26. Hamby, A., Ecker, U. K. H., Brinberg, D. (2020). How stories in memory perpetuate the continued influence of false information. *Journal of Consumer Psychology* 30(2), 240–259.
27. Johnson, H. M., Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(6), 1420–1436.

28. Jones, K. A., Crozier, W. E., Strange, D. (2017). Believing is seeing: Biased viewing of body-worn camera footage. *Journal of Applied Research in Memory and Cognition*, 6(4), 460–474.
29. Kassin, S. M., Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the “harmless error” rule. *Law and Human Behavior*, 21(1), 27–46.
30. Kramer, G. P., Kerr, N. L., Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior*, 14, 409–438.
31. Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Slovic, 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.
32. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., Cook, J. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131.
33. Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science*, 16(3), 190–195.
34. Lewandowsky, S., Stritzke, W., Oberauer, K., Morales, M. (2009). Misinformation and the “War on terror”: when memory turns fiction into fact. (In) W. G. K. Stritzke, S. Lewandowsky, D. Denemark, J. Clare, F. Morgan (Eds.), *Terrorism and torture an interdisciplinary perspective* (pp. 179–203). Cambridge University Press.
35. Lindsay, D. S., Johnson, M. K. (1989a). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17, 349–358.
36. Lindsay, D. S., Johnson, M. K. (1989b). The reversed eyewitness suggestibility effect. *Bulletin of the Psychonomic Society*, 27, 111–113.
37. Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7(4), 560–572.
38. Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366.
39. 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.
40. McCloskey, M., Zaragoza, M. (1985). Misleading post-event information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114(1), 1–16.
41. Memon, A., Meissner, C. A., Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16(4), 340–372.
42. Misinformation. (In) *Merriam-Webster’s collegiate dictionary*. Retrieved September 7, 2020 from: <https://www.merriam-webster.com/dictionary/misinformation>.
43. Mitchell, K. J., Johnson, M. K. (2000). Source monitoring: Attributing memories to sources. (In) E. Tulving, F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). New York: Oxford University Press.
44. Pansky, A., Tenenboim, E., Bar, S. K. (2011). The misinformation effect revisited: Interactions between spontaneous memory processes and misleading suggestions. *Journal of Memory and Language*, 64(3), 270–287.
45. Polak, M., Dukała, K., Szpitalak, M., Polczyk, R. (2016). Toward a non-memory misinformation effect: Accessing the original source does not prevent yielding to misinformation. *Current Psychology*, 35(1), 1–12.
46. Poland, G. A., Spier, R. (2010). Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine*, 28, 2361–2362.
47. Polczyk, R. (2007). *Mechanizmy efektu dezinformacji w kontekście zeznań świadka naocznego*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
48. Polczyk, R. (2017). The “memory” misinformation effect may not be caused by memory failures: Exploring memory states of misinformed subjects. *Polish Psychological Bulletin*, 48(3), 388–400.
49. Prokhorov, A. M. (Ed.). (1974–1983). *Great Soviet Encyclopedia. Translation of third Russian edition of Bol’shaya sovetskaya entsiklopediya*. New York: Macmillan Publishers.
50. Rajczyk, R. (2016). *Nowoczesne wojny informacyjne*. Warszawa: Difin.
51. Steblay, N., Hosch, H. M., Culhane, S. E., McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior*, 30, 469–492.
52. Swire, B., Berinsky, A. J., Lewandowsky, S., Ecker, U. K. H. (2017). Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science*, 4(3), 160802.
53. Swire, B., Ecker, U. K. H., Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1948–1961.
54. Swire-Thompson, B., Ecker, U. K. H., Lewandowsky, S., Berinsky, A. J. (2019). They might be a liar but they’re my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, 41(1), 21–34.
55. Szpitalak, M. (2012). *Motywacyjne mechanizmy efektu dezinformacji*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
56. Szpitalak, M. (2015). *W kierunku poprawy jakości zeznań świadków: pozytywne i negatywne następstwa ostrzeżenia o dezinformacji*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

57. Szpitalak, M., Polczyk, R. (2011). Efekt rozgrzania poznawczego a podatność na dezinformację. *Psychologia Społeczna*, 3, 214–225.
58. Szpitalak, M., Polczyk, R. (2012). Efekt wzmocnionej autoafirmacji jako metoda redukcji efektu dezinformacji. (In) D. Doliński, J. Maciuszek, R. Polczyk (Eds.), *Wokół wpływu społecznego* (pp. 99–111). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
59. Szpitalak, M., Polczyk, R. (2017). Efekt dezinformacji z perspektywy psychologii społecznej: natura i uodparnianie. *Psychologia Społeczna*, 12(1), 30–41.
60. Szpitalak, M., Polczyk, R. (2019). How to induce resistance to the misinformation effect? Characteristics of positive feedback in the reinforced self-affirmation procedure. *Psychology, Crime & Law*, 25(7), 771–791.
61. Wachowicz, M. J. (2019). Ujęcie teoretyczne pojęcia dezinformacji. *Wiedza Obronna*, 226–253.
62. Walter, N., Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, 47(2), 155–177.
63. Wilkes, A. L., Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 40A(2), 361–387.
64. Wright, D. B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory*, 1(2), 153–166.

Corresponding author

Assoc. Prof. Malwina Szpitalak
Jagiellonian University
Institute of Psychology
ul. Ingardena 6
PL 30-060 Kraków
e-mail: malwina.szpitalak@uj.edu.pl

EFEKT PRZEDŁUŻONEGO WPŁYWU DEZINFORMACJI W KONTEKŚCIE PSYCHOLOGII SĄDOWEJ

Dezinformacja – geneza pojęcia i definicje

Termin „dezinformacja” ma najprawdopodobniej pochodzenie rosyjskie (*дезинформация; dezinformatsiya*) i w tym rozumieniu dotyczy celowego rozpowszechniania fałszywych informacji z zamiarem wywarcia wpływu na postawy tych, do których te informacje docierają (Prokhorov, 1974–1983). W języku angielskim istnieje również znacznie starsze, datowane na XVI wiek (Online Etymology Dictionary, b.d.) pojęcie *misinformation*, oznaczające nieprawdziwą lub niedokładną informację, która, w odróżnieniu od *disinformation*, nie jest rozpowszechniana celowo (por. „disinformation”, Online Etymology Dictionary, b.d.; „misinformation”, Merriam-Webster’s collegiate dictionary, b.d.). W języku polskim brak tego rozróżnienia. Ten z pozoru błahy problem może prowadzić do niepełnego lub nieprawidłowego obrazu zjawiska dezinformacji w dziedzinach innych niż te najczęściej wymieniane przez rodzime piśmiennictwo, w którym dezinformację traktuje się np. jako formę przekazu wprowadzającą odbiorcę w błąd i ingerencję w jego proces decyzyjny (Basaj, 2018), część składową propagandy (Rajczyk, 2016) lub mylne informowanie w sytuacji braku rzetelnych informacji (Dubisz, 2003), co odnosi się zarówno do mediów, środowiska internetowego i mediów społecznościowych (*fake-news*; Brzóska, Nowak, 2019), jak i podmiotów w obrębie środowiska militarnego, ekonomicznego, naukowo-technologicznego i politycznego (Wachowicz, 2019). W przeglądach tych konsekwentnie nie podejmuje się omawiania dezinformacji w rozumieniu psychologicznym, co może być m.in. konsekwencją stosowania w literaturze zagranicznej raczej wyrażenia *misinformation* zamiast *disinformation*. W polskiej literaturze psychologicznej można z kolei odnaleźć informacje o efekcie dezinformacji, definiowanym jako włączanie przez osobę dokonującą raportu pamięciowego informacji niezgodnych z przebiegiem wydarzenia i pochodzących z innych niż ono samo źródeł (Polczyk, 2007). W literaturze anglojęzycznej, oprócz wspomnianego efektu dezinformacji, badacze zajmują się również tzw. efektem przedłużonego wpływu (*continued influence effect*; Johnson, Seifert, 1994) – ciągłym wpływem pewnej informacji na relacje i wnioski o zdarzeniu. Co istotne, informacja ta, mimo że została skorygowana i uznana za błędną, wciąż jest powielana i wpływa na sądy osoby, która miała z nią styczność. Choć jest to temat chętnie eksplorowany, w polskiej literaturze przedmiotu brakuje eksperymentalnych badań w tej dziedzinie. Niniejszy artykuł ma na celu przedsta-

wienie i porównanie obydwu wyżej wymienionych zjawisk.

Efekt dezinformacji (*post-event misinformation effect; PME*)

Efekt dezinformacji rozpatruje się zwykle w kategoriach rekonstruktywnego charakteru pamięci (Mitchell, Johnson, 2000). W badaniach nad nim zwraca się najczęściej uwagę na kontekst sądowy (Loftus, 2005) – świadek jest narażony na odbiór informacji docierających do niego z różnych źródeł, innych niż samo zdarzenie. Jeżeli informacje te są nieprawdziwe, raport pamięciowy świadka – mimo że niecelowo i w dobrej wierze – może być zniekształcony i mieć konsekwencje dla postępowania sądowego.

Charakterystyczny dla PME jest jego trzyetapowy przebieg, obecny zarówno w warunkach nielaboratoryjnych, jak i odwzorowywany w procedurach badawczych (np. Loftus, Miller, Burns, 1978). W pierwszym etapie badani zapoznają się z materiałem oryginalnym (może to być np. fragment filmu, tekst czytany lub inscenizacja), a po jakimś czasie – z materiałem następczym (np. opisem wydarzenia lub relacją innej osoby), który zawiera dezinformację. W ostatnim etapie badani odpowiadają na pytania dotyczące materiału oryginalnego.

Uzyskiwane w podobnych jak opisane powyżej warunkach eksperymentalnych różnice między grupami dezinformowanymi i niedezinformowanymi osiągają zwykle wartość kilkudziesięciu procent i są doskonale replikowalne (por. Polczyk, 2007).

Mechanizmy efektu dezinformacji

Choć fakt istnienia PME z reguły nie podlega dyskusji, bardziej problematyczną kwestią pozostają jego mechanizmy. Na początku badań w tej dziedzinie przyjmowano niemal wyłącznie koncepcje dotyczące błędów funkcjonowania systemu pamięciowego (Loftus i in., 1978). Najwcześniejsza z nich głosiła, że ślad pamięciowy informacji oryginalnej jest wskutek obecności dezinformacji uszkodzony (Loftus, 1975) lub nawet nadpiswany (Loftus i in., 1978). W późniejszych teoriach skupiano się na współlistnieniu obydwu śladów pamięciowych, z których ten zawierający dezinformację pogarsza dostęp do oryginalnego (Bekerian, Bowers, 1983), szybszej dezintegracji reprezentacji formy wydarzenia przy

zachowaniu jej treści, która może być przez to generalizowana na szczegóły dezinformowane (Brainerd, Reyna, 2002) lub współzawodnictwie asocjacji i błędów identyfikacji ich źródła (w konsekwencji błędnego inferowania źródła aktywacji określonej reprezentacji; Ayers, Reder, 1998). Wydaje się jednak, że najbardziej uznana i współcześnie dominująca jest teoria błędów w monitorowaniu źródła informacji (Lindsay, Johnson, 1989a), w której za przyczynę PME uznaje się nieprawidłowe przypisywanie źródła informacji do jej treści. Teoria ta jednak nie okazała się kompletna (Polczyk, 2007). Istnieje szereg badań wskazujących na to, że raport pamięciowy może być niezgodny z materiałem oryginalnym również wtedy, gdy pamięć materiału oryginalnego pozostaje nienaruszona (np. Blank, 1998; Polak, Dukala, Szpitalak, Polczyk, 2016; Polczyk, 2007; 2017). W związku z tym niektórzy badacze sugerują, by zjawisko pamięci szczegółów i pamięci źródła uznać za dwa osobne przejawy PME (Pansky, Tenenboim, Bar, 2011).

Pierwsze badania podważające wyłącznie pamięciowe mechanizmy efektu dezinformacji opublikowali McCloskey i Zaragoza (1985), jednak to badania Blanka (1998) zapoczątkowały systematyczną eksplorację mechanizmów pozapamięciowych. Dzięki zastosowaniu procedury czterostopowej polegającej na dodaniu drugiego testu końcowego, poprzedzonego poinformowaniem badanych co do istnienia rozbieżności między materiałem oryginalnym a następczym, Blank odkrył, że średnio około 40% osób badanych, które uległy dezinformacji, było świadomych istnienia rozbieżności – wyniki te tłumaczono głównie niepewnością własnej pamięci badanych.

Efekt przedłużonego wpływu (*continued influence effect; CIE*)

W obliczu powszechnego wpływu dezinformacji na różne aspekty funkcjonowania jednostek i większych grup społecznych istnieje potrzeba skutecznego jej korygowania. Zadanie to jest o tyle trudne, że raz przyswojona dezinformacja nie daje się łatwo usunąć. Pionierami badania tego zjawiska, określanego jako efekt przedłużonego wpływu, byli Johnson i Seifert (1994) oraz Wilkes i Leatherbarrow (1988). Klasyczny paradygmat badania CIE obejmuje przedstawienie uczestnikom scenariusza fikcyjnego zdarzenia, a następnie tzw. wycofanie informacji o przyczynie tego incydentu, tj. poinformowanie badanego, że w rzeczywistości informacja ta nie była prawdziwa. Na przykład badanym prezentowano fikcyjne doniesienia na temat pożaru lokalu, w którym znajdowały się materiały papiernicze. Pierwsza wersja zawierała informację, że w pomieszczeniu, w którym miało miejsce zdarzenie, przechowywano materiały łatwopalne bez dostatecznego zabezpieczenia. Następnie demento-

wano tę informację, podając, że takich materiałów tam nie było. Mimo pamięci o wycofaniu informacji badani w końcowym teście nadal formułowali wnioski, które zawierały informacje o materiałach łatwopalnych. Efekt ten jest dobrze replikowalny i w większości badań znaczna część badanych posługuje się dezinformacją w swoich sądach (np. Ecker, Ang, 2019; Ecker, Hogan, Lewandowski, 2017; Ecker, Lewandowski, Swire, Chang, 2011).

Mechanizmy efektu przedłużonego wpływu

W literaturze można znaleźć dwa dominujące, choć niewykluczające się poznawcze wyjaśnienia mechanizmów CIE. Autorzy pierwszego sugerują, że jest on wynikiem błędu podczas wyszukiwania informacji (np. Ecker i in., 2010) i zakładają, że dezinformacja i jej odwołanie przechowywane są w obrębie różnych reprezentacji w pamięci i że rywalizują one o aktywację podczas wyszukiwania. Dezinformacja wpływa na wnioskowanie wtedy, kiedy to ona, a nie korekta, zostaje selektywnie wybrana. Koncepcja ta łączy działanie zarówno automatycznych, jak i strategicznych procesów pobierania i monitorowania: gdy dezinformacja jest aktywowana automatycznie za pomocą wskazówek, wymagane są procesy strategiczne, aby ocenić ważność uzyskanych informacji i/lub przypomnieć sobie wszelkie istotne poprawki. Ponieważ procesy strategiczne są poznawczo wymagające, często zawodzą, co skutkuje efektem przedłużonego wpływu.

Alternatywnym wyjaśnieniem jest teoria modeli mentalnych (Ecker i in., 2011) opierająca się na sposobie, w jaki ludzie przetwarzają informacje o wydarzeniach. Modele mentalne, budowane na podstawie pozyskiwanych treści, składają się z elementów połączonych ze sobą w łańcuch przyczynowo-skutkowy. Wycofanie informacji z powstałego modelu powoduje powstanie luki, która czyni go mniej kompletnym. Dezinformacja, choć została wycofana, może tę lukę wypełnić, pozwalając na zbudowanie modelu kompletnego, choć potencjalnie nieprawidłowego, który jest jednak lepszy od dokładnego, ale niekompletnego (Ecker i in., 2011; Walter, Tukachinsky, 2020). Za teorią modeli mentalnych przemawiają pewne behawioralne i neuropsychologiczne dowody: badania neuroobrazowe z użyciem fMRI wykazały, że aktywność neuronów związanych z integracją informacji jest różna w sytuacji wycofania dezinformacji lub jego braku, co może sugerować, że niepowodzenie w integracji i aktualizacji pamięci stoi za CIE (Gordon, Brooks, Quadflieg, Ecker, Lewandowski, 2017). Inne badania pokazują, że istnieje większe prawdopodobieństwo niepowodzenia integracji, gdy osoba nie wykorzystuje w pełni swoich zasobów poznawczych np. z powodu obciążenia poznawczego przy jednoczesnym kodowaniu wycofania informacji (Ecker i in., 2011).

Hamby, Ecker i Brinberg (2020) stwierdzają, że nie-skuteczność wycofywania dezinformacji ma związek z relacją, jaka zachodzi między kluczową informacją a szerszym kontekstem (tym, co człowiek miał w pamięci wcześniej) i że wycofane dezinformacje mają silniejszy wpływ na rozumowanie i przekonania jednostki, jeśli mogą retrospektywnie wyjaśnić historię, którą wcześniej miała ona w pamięci. Uczestnikom eksperymentu przedstawiono historię o osobie, która doświadczyła komplikacji po zażyciu pewnego specyfiku. Część badanych otrzymała w wyjaśnieniach powód tych komplikacji, a inną grupę pozostawiono bez wyjaśnień. Następnie obie grupy otrzymywały inną informację, która potencjalnie mogła wyjaśnić te komplikacje, po czym informację tę wycofywano. Podczas testu końcowego okazało się, że osoby pozbawione wstępnej informacji o powodach komplikacji dalej ulegały dezinformacji, używając jej do wnioskowania. Ta błędna informacja pomogła wypełnić lukę w ich modelu mentalnym. Badani, którzy na początku otrzymywali wyjaśnienie o powodach komplikacji, nie mając luki do wypełnienia, rzadziej polegali na wycofanej dezinformacji.

Wyznaczniki efektu przedłużonego wpływu

Choć czynniki poznawcze odgrywają znaczącą rolę w działaniu efektu przedłużonego wpływu, nie mogą całkowicie go wyjaśnić. W literaturze pod uwagę brane są także czynniki motywacyjne, takie jak postawy lub światopogląd. Warto zauważyć, że duża część badań dotyczących CIE odnosi się do tego, jak działa on na rozumowanie i sądy jednostki, mniej uwagi poświęcając faktycznemu wpływowi na zachowania i wybory ludzi. Wyniki badań (Swire, Berinsky, Lewandowski, Ecker, 2017) pokazują zaś, że korygowanie dezinformacji niekoniecznie przekłada się na zmianę zachowań – ludzie nie zmieniają zamiaru głosowania na polityków, wobec których zmodyfikowali swój osąd wskutek weryfikacji ich wypowiedzi czy działań (np. kłamstw; Swire, Ecker, Lewandowski, Berinsky, 2020). Ecker i Ang (2019), badając zwolenników prawicowych i lewicowych partii w Australii, stwierdzili, że postawy stronnicze mają wpływ na przetwarzanie wycofania, szczególnie jeśli dezinformacja dotyczy ogólnych poglądów, odnoszących się do danego skrzydła politycznego – wycofanie dezinformacji jest bardziej skuteczne, jeśli nie wymaga zmiany nastawienia politycznego. Szczególnie trudno jest wycofać dezinformację, jeśli jest ona zgodna z poglądami konserwatywnymi. Korekty niezgodne ze światopoglądem mogą nawet, w określonych warunkach, przynieść efekt odwrotny do zamierzonego (*backfire*) czyli paradoksalnie zwiększyć uleganie dezinformacji (Lewandowski, Ecker, Seifert, Schwarz, Cook, 2012; wyniki przeciwne: Ecker i in., 2014).

Rola światopoglądu, osobistych sympatii czy uprzedzeń nie ogranicza się tylko do ulegania dezinformacji w kwestiach politycznych. W badaniu Eckera i in. (2014) uczestnikom o wysokim i niskim poziomie uprzedzeń rasowych przedstawiono wydarzenie dotyczące fikcyjnego napadu rabunkowego, w którym rdzenny Australijczyk w jednej wersji był podejrzanym o ten napad, a w innej był bohaterem, który napad udaremnił. Okazało się, że poglądy rasowe nie determinują skuteczności wycofania dezinformacji, choć dało się zauważyć, że osoby o wysokim stopniu uprzedzeń rasowych częściej w swoich wypowiedziach wspominały o Aborygenie podejrzanym niż o Aborygenie bohaterze. Podobne wyniki uzyskano w badaniu, w którym fikcyjną katastrofę samolotu przypisano najpierw atakowi terrorystycznemu, a następnie wprowadzono korektę, wyjaśniając, że samolot miał uszkodzony zbiornik paliwa. Osoby z wysokim poziomem islamofobii w teście wnioskowania częściej nawiązywały do wątków terrorystycznych niż osoby z jej niskim poziomem, choć wycofanie dezinformacji było jednakowo skuteczne w obu grupach. Co ciekawe, w badaniu tym stwierdzono także, że emocjonalność historii nie ma wpływu na stopień ulegania dezinformacji, a informacje pełne emocji nie są bardziej odporne na korektę niż te o mniejszym ładunku emocjonalnym (Ecker, Lewandowski, Apai, 2011).

Czynnikiem potencjalnie zmniejszającym CIE jest podejrzliwość czy to wobec źródła informacji, czy intencji uczestników danego zdarzenia. Badania dotyczące wojny w Iraku w 2003 roku prowadzone w trzech grupach uczestników, z których dwie pochodziły z krajów koalicyjnych (USA i Australia), a trzecia z przeciwnych wojnie Niemiec, pokazały, że uczestnicy z USA byli najmniej wrażliwi na korektę dezinformacji związanych z tym wydarzeniem. Niemcy, jako najbardziej sceptyczni odnośnie do oficjalnego powodu tej wojny (zniszczenie broni masowego rażenia), rzadziej ulegali dezinformacji (Lewandowski i in., 2005). Niemniej jednak w wielu sytuacjach wzbudzenie podejrzliwości może nie być możliwe, jest ona bowiem związana ze sceptycyzmem, który jest stabilną cechą osobowości (Lewandowski, Stritzke, Oberauer, Morales, 2005; 2009).

Analizowane jest też powtarzanie dezinformacji w komunikacji wycofującym ją (Ecker i in., 2017, 2011). Wśród badaczy trwa debata na temat tego, czy błędna informacja powinna zostać powtórzona w sprostowaniu. Dość powszechny jest pogląd, że powtarzanie dezinformacji podczas jej wycofania może spowodować tzw. efekt zaznajomienia polegający na tym, że coś, co brzmi znajomo, prawdopodobnie częściej zostanie uznane za prawdę (Swire, Ecker, Lewandowski, 2017) – w kontekście CIE powtarzana w wycofaniu dezinformacja mogłaby w przyszłości łatwiej przychodzić na myśl, np. wyjaśnienie, że szczepionki nie powodują autyzmu, zawiera w sobie skojarzenie słów „szczepionka” i „autyzm”, co

sprawia, że fałszywe stwierdzenie o negatywnym wpływie szczepionek będzie brzmiało znajomo (Lewandowski i in., 2012; Swire i in., 2017). Wyniki części badań wskazują jednak na efekty przeciwne (Ecker i in., 2017), sugerując, że powtarzanie dezinformacji w komunikacie, który ją odwołuje, można uznać za skuteczną i bezpieczną strategię. Jednocześnie badacze wyrazili pewne zastrzeżenia: po pierwsze – odstęp czasu między przedstawieniem błędnych informacji a ich wycofaniem był krótszy niż miałyby to miejsce w wielu rzeczywistych sytuacjach (czas z kolei może wpływać na skuteczność wycofania; Walter, Tukachinsky, 2020); po drugie – wycofania zawierające przyczyny alternatywne w prawdziwym świecie często nie są dostępne. Jest możliwe, że powtarzanie dezinformacji może mieć negatywny wpływ tylko w przypadku braku takich alternatyw (Ecker i in., 2017), jednak Swire i in. (2017) nie znaleźli na to żadnego dowodu. Ecker, O'Reilly, Reid i Chang (2020) sugerują natomiast, że w przypadku kiedy badani zapoznają się z nieznaną im wcześniej dezinformacją, może jednak wystąpić efekt zaznajomienia i związane z tym niebezpieczeństwo rozprzestrzeniania się jej w populacji, lecz wyniki innych badań (Ecker i in., 2020) pokazują, że nawet kiedy dezinformacje nie były wcześniej znane, ich powtarzanie w komunikacie wycofującym nie powodowało zwiększenia efektu przedłużonego wpływu.

Techniki uodparniania

Bardzo ważnym zagadnieniem z obszaru badań nad CIE z perspektywy praktycznej są metody uodparniania na wpływ dezinformacji. Konsekwencje ulegania fałszywym informacjom rozpowszechnianym w przestrzeni publicznej widać chociażby na przykładzie skutków *fake-newsów* (Lazer i in., 2018) lub wiary w to, że szczepionki są przyczyną autyzmu (Poland, Spier, 2010). Przykłady te dobrze obrazują problem, że zwykle wycofanie lub korekta fałszywych informacji nie powstrzymuje ich wpływu na poglądy jednostki. Jak dotąd zidentyfikowano trzy czynniki redukujące wpływ dezinformacji w paradygmacie CIE: ostrzeżenie, powtórzenie wycofania i przedstawienie alternatywy.

Przedstawienie alternatywy – jedną z wyżej wymienionych technik jest, bazujące na wspomnianej wcześniej teorii modeli mentalnych, przedstawienie alternatywnej przyczyny zdarzenia, zastępującej dotychczasową dezinformację (Ecker i in., 2010; Johnson, Seifert, 1994) i uwypuklającą sprzeczność między słusznymi i nieuprawnionymi informacjami, co znacząco ułatwia procesy integracji/aktualizacji. Uważa się, że ludzie preferują zachowanie w pamięci informacji o znaczeniu kluczowym, czyli takich, które dotyczą przyczyn zdarzeń oraz zaangażowanych w nie osób, a podanie innego niż dezinformacja wyjaśnienia wspomaga proces właściwe-

go wnioskowania, jednocześnie uzupełniając lukę w modelu mentalnym bez szkody dla jego wiarygodności. Ecker i in. (2011) manipulowali również siłą wycofywania dezinformacji przez kilkukrotne powtarzanie informacji korygujących, co, ich zdaniem, może prowadzić do silniejszej aktualizacji mentalnego modelu bezpośrednio po zakodowaniu wycofania. Chociaż technika obalenia i powtórzonego wycofania są bardziej skuteczne niż samo wycofanie, redukcja CIE nigdy nie jest całkowita.

Ostrzeżenie przed możliwością wystąpienia dezinformacji dość szeroko przebadano w paradygmacie PME (np. Blank, Launay, 2014; Chambers, Zaragoza, 2001; Greene, Flynn, Loftus, 1982; Szpitalak, 2015; Wright, 1993). Jest ono techniką prostą w indukcji, trafną ekologicznie i skuteczną niezależnie od umiejscowienia (por. Szpitalak, 2015). Natomiast prawdopodobnie pierwsze badanie, w którym zastosowano ostrzeżenie o dezinformacji w paradygmacie efektu przedłużonego wpływu, przeprowadzili Ecker i in. (2010). W swoich eksperymentach prezentowali badanym fikcyjną relację z wypadku minibusa, którym według pierwszego doniesienia podróżowali starsi podopieczni domu opieki. W następnych etapach w zależności od grupy (1) informację wycofywano lub (2) zastępowano inną (podając, że pasażerami byli zawodnicy studenckiej drużyny hokejowej), (3) wycofywano oraz dawano wcześniej pisemne ostrzeżenie o tym, że czasem pierwsze doniesienia z wydarzeń mogą być niedokładnie sprawdzone i nieprawdziwe i (4) informowano o istnieniu CIE z konkretnymi przykładami jego działania. Ogólne ostrzeżenie w warunku (3) okazało się skuteczne tylko w przypadku tych badanych, którzy pamiętali o wycofaniu dezinformacji, natomiast manipulacja w warunku (4) zmniejszyła zależność od dezinformacji w stopniu podobnym do podania wiarygodnej alternatywy. W kolejnym eksperymencie połączono dwie najskuteczniejsze manipulacje, czyli podanie wiarygodnej alternatywy z jednoczesnym ostrzeżeniem o CIE. Uzyskano obniżenie ulegania dezinformacji w stopniu większym niż przy samych technikach składowych, niemniej jednak efekt nie został całkowicie wyeliminowany (Ecker i in., 2010).

Implikacje praktyczne i dalsze kierunki badań

Pomimo ponad czterdziestu lat badań psychologicznych nad omawianymi efektami dezinformacji istnieje bardzo niewiele prób ich integracji (por. Ecker i in., 2015). Wydaje się to dość zaskakujące, zwłaszcza biorąc pod uwagę podobieństwa między eksperymentami tradycyjnie zaliczanymi do nurtu PME a zjawiskiem CIE (np. Lindsay, Johnson, 1989b).

Jedną z korzyści wspólnego badania PME i CIE byłaby możliwość głębszego zrozumienia obydwu zjawisk i utworzenie skuteczniejszych technik uodparniania,

zwłaszcza że z tych dotychczas opracowanych wynika, że nadają się one do zastosowania w warunkach pozalaboratoryjnych w różnym stopniu. W przypadku PME wiele z nich ma raczej charakter podstawowy (choć część wykazuje się wysoką skutecznością i możliwością zastosowania w praktyce, np. Greene i in., 1983; Memon, Meissner, Fraser, 2010; Szpitalak, 2012; Szpitalak, Polczyk, 2012; 2019), z kolei repertuar metod uodparniających na CIE wydaje się bardzo ubogi i niezbyt dobrze zbadany. Teoretyczne spekulacje odnośnie do wykorzystania technik uodparniania na PME wobec CIE wydają się obiecujące: jak zauważają Szpitalak i Polczyk (2017), bardzo ważnym krokiem w stronę ulepszenia metod uodparniania byłoby sprawdzenie ich w przypadku, gdy osoba jest świadoma niezgodności między docierającymi do niej informacjami, co zostało dotychczas przeprowadzone dla ostrzeżenia (Polczyk, 2007) i wzmocnionej autoafirmacji (Szpitalak, 2012), wskazując na większą efektywność metod właśnie w przypadku, gdy osoba posiada wspomnienia zarówno informacji oryginalnej, jak i dezinformacji – w przypadku ostrzeżenia przed CIE wygląda to identycznie (Ecker i in., 2010).

Dotychczasowa wiedza oraz nowe badania dotyczące CIE mogłyby mieć natomiast spore znaczenie dla stosowanej psychologii sądowej. Informacje, które bywają często ujawniane w mediach przed rozprawami, zagrażają bezstronności wymiaru sprawiedliwości, dlatego tak ważne jest unikanie ujawniania określonych kategorii informacji przed rozpoczęciem procesu. Gdy wprowadzają one wątpliwości co do niewinności oskarżonego lub przedstawiają go w negatywnym świetle, może to skutkować częstszymi wyrokami uznającymi winę oskarżonego. Np. Gilbert, Tatarodi i Malone (1993), prezentując badanym fałszywe, zaostrzające winę informacje na temat osoby oskarżonej o przestępstwo, wykazali, że pomimo czytania przez uczestników eksperymentu informacji prawdziwych, przeczących winie oskarżonego, oceniali oni, że oskarżony zasługuje na wyższą karę pozbawienia wolności. Black (2018) z kolei sugeruje istotność uodparniania na CIE w przypadku wycofywania zeznań świadków oraz uprzedzeń sędziów: jak wspomniano przy omawianiu PME, zeznania naocznych świadków mogą być zniekształcone, a potem, pomimo ich wycofywania, nadal wpływać na decyzje wymiaru sprawiedliwości, zwłaszcza jeśli jego przedstawiciele byłiby mało profesjonalni, stronniczy czy uprzedzeni. Autorka proponuje również, by powstrzymać się od informowania o wycofywaniu zeznań do czasu uzyskania alternatywnych wyjaśnień sprawy. Można również połączyć CIE z problemem fałszywych samooskarżeń: sędziowie, zapoznając się z zeznaniami osób, które w trakcie przesłuchań były zmuszane do przyznania się do winy, wydawali skazujące wyroki tak samo często, jak wobec osób przyznających się bez przymusu. Działo się to pomimo deklaracji sędziów, że fakt, iż zeznania były wymuszono-

ne, nie będzie wpływać na ich decyzje (Kassin, Sukel, 1997). Jones, Crozier i Strange (2017) z kolei, łącząc paradygmat PME i CIE, przedstawili wyniki świadczące o tym, że nawet przy dowodach bezpośrednich ludzie są w stanie posługiwać się wnioskami zgodnymi z dezinformacją. Badani zapoznawali się z opisem sprzeczki między obywatelem a funkcjonariuszem policji oraz tym, jak się ona skończyła, a następnie czytali zawierający dezinformację raport policyjny na temat tego zdarzenia (obwiniający obywatela i usprawiedliwiający policjanta) oraz oglądali nagranie wideo z noszonej przez tego samego funkcjonariusza kamery (*body worn camera; BWC*). Okazało się, że pomimo obejrzenia bezpośredniego zapisu wydarzenia badani generowali wnioski zgodne z policyjną dezinformacją, zrzucając winę na obywatela.

Przysięgli przed uzgodnieniem wyroku otrzymują od sędziego instrukcję mówiącą o tym, które dowody pojawiające się w śledztwie nie zostały dopuszczone do rozprawy sądowej (w efekcie przedłużonego wpływu odpowiadałoby to wycofaniu nieprawdziwych informacji, Fein, McCloskey, Tomlinson, 1997). Często ignorują oni takie instrukcje i orzekają, uwzględniając te wycofane dowody (jest to przejaw efektu przedłużonego wpływu). Steblay, Hosch, Culhane i McWethy (2006) donoszą na podstawie swoich badań, że jeśli sędzia w swojej instrukcji uzasadni przysięgłym, dlaczego dany dowód został wycofany (odpowiednik wiarygodnej alternatywnej informacji), to sędziowie starannie wypełniają instrukcje (odpowiednik zmniejszenia efektu przedłużonego wpływu).

Jednym ze środków zaradczych przeciwko CIE bywa stosowane przez sądy odroczenie rozpoczęcia procesu. Jeśli przedprocesowe relacje medialne na temat sprawy przedstawiają informacje sugerujące winę oskarżonego, obrona może wystąpić o takie rozwiązanie w nadziei, że szum medialny ucichnie. Wyniki badań sugerują, że takie opóźnienie może mieć pozytywne skutki (Kramer, Kerr, Carroll, 1990). Ecker i in. (2010) uważają natomiast, że ostrzeżenia wydają się obiecującym i ekonomicznym sposobem na zapobieganie sytuacjom, w których sędziowie polegają na nieaktualnych już informacjach przy wydawaniu osądów, i sugerują, by konkretne ostrzeżenia podawać sędziom rutynowo z góry, aby wzbudzić ich podejrzenia odnośnie do trafiających do nich informacji. Podobne wnioski można wyciągnąć ze starszych badań – Fein i in. (1997) wykazali, że wpływ wycofanych fałszywych dowodów i oskarżeń na sędziów można zredukować, jeśli przed ich prezentacją wzbudziło się w badanych podejrzenia co do powodów, dla których te informacje zostały udostępnione.

W przypadku zeznań świadków schemat CIE wydaje się dopełniający wobec PME: gdy świadek zdarzenia napotyka na dezinformację, która jest następnie wycofywana, może nadal polegać na niej w swoich zeznaniach. Przypomina to znacząco używaną w nurcie PME proce-

durę zanegowanej dezinformacji (Lindsay, 1990), w której osobę badaną informuje się, że jeden ze szczegółów z materiału oryginalnego został opisany błędnie oraz że materiał następczy nie zawiera poprawnych odpowiedzi na pytania testu końcowego. Tym samym wartość dotychczasowej wiedzy o CIE w przypadku zeznań świadków może się okazać nieoceniona: stosowanie procedur ostrzegania, obalania lub wielokrotnie powtarzanego wycofania oraz potencjalnych nowych technik mogłoby zmniejszać szansę na uleganie dezinformacji w przypadku, gdy ktoś nie był bezpośrednim świadkiem wydarzenia lub szczególnie jeśli nie zapamiętał jego przebiegu, a mimo to zeznaje, kierując się informacjami wywiedzionymi z doniesień prasowych, plotek lub rozmów z innymi osobami.