PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Perceived memory credibility: The role of details

Lynn Nadel[a,b,1] and Katharine C. Simon[c,d,1]

Contributed by Lynn Nadel; received August 14, 2024; accepted November 19, 2024; reviewed by Henry L. Roediger III and John T. Wixted

The sharing of personal memories is a unique aspect of the human experience. Humans communicate to provide information, to influence, or even to amuse. How do we distinguish between credible and noncredible narratives? Forensic science has identified race, age, and detail quantity as influential. We do not know how the nature of narrated details impacts believability. We report two studies investigating how detail composition influences credibility ratings using definitions of details suggested by Levine et al. (2002). Internal details are directly connected to the episodic aspects of the memory, while external details refer to semantic facts or depictions not directly related to the main event. A total of 825 participants rated narratives that varied detail number and type for perceived credibility or saliency. Episodic memory details enhanced credibility more than semantic memory details. In addition, within episodic memories, person-related details enhanced credibility more than non-person-related details. Our results suggest a lens through which to view the credibility of what we hear and read.

episodic memory | believability | credibility

Significance

Outside the forensic domain, little is known about the factors that influence credibility judgments in humans. Our research focuses on the details included within memory narratives and highlights the role various types of detail play in reported credibility and salience. Memory narratives that included episode-specific details were rated as more credible than those with general details.

The field of autobiographical episodic memory research is expansive, largely focusing on determining how consciously formed memories are encoded, stored, and retrieved and the neural structures involved (see reviews, 1, 2). Retained episodic memories serve many purposes: They can alter personal behavior, justify beliefs, values, or judgments of others, or be used in social interactions (3–5). Sharing memories with others is adaptive, the goal of which may be to signal value alignment, impact others' emotional states, empathize, influence others' choices, or propagate information forward (see 3). Information, however, is useful only if the story conveying it is believable, whether it refers to an actual event or is a part of a fictional storyline. How the composition of details in a shared memory influences its believability is poorly understood.

There is surprisingly sparse experimental literature on memory believability, much of it generated within a forensic framework regarding eyewitness testimony, although interest in what it takes to create a believable artificial agent has emerged in recent years. One notable exception is work carried out within Johnson's "source monitoring" framework (e.g., 6). According to this framework, individuals monitor the source of their own retrieved memories, distinguishing between events that actually transpired from those that they only thought about. Johnson et al. (7) asked whether this logic could also be applied to evaluating the reports of others (7). Participants in their study were instructed to say whether they believed people's reported memories had been experienced or imagined instead. The "memories" were accounts of everyday experiences, which varied in the extent of both perceptual and emotional details. Johnson et al. found that believability increased with added detail and that both perceptual and emotional details contributed to this effect. Using similar methods, Sporer & Sharman (8) supported these findings, as did Wright et al. (9). Details, in other words, are important (8, 9).

Some two decades ago Levine and colleagues described a now widely adopted Autobiographical Memory Interview methodology predicated on the notion that memory details can be separated into two broad categories—internal and external details (10). The former are specific details that connect to the episodic nature of the story, such as the time and place of occurrence, while the latter are semantic in nature, with no reference to a specific time and place, including editorializing and metacognitive statements (10). This distinction resembles but does not directly map onto, how Johnson et al. separated perceptual from emotional memory details in their study of believability. Further, this distinction expands beyond the episodic-semantic dichotomy with the additional editorializing and metacognitive statements. The possibility that an "internal vs external" framing captures the organization of the neurocognitive systems underlying memory motivated our focus on the impact of such details on believability.

More recently, solving how to gauge a memory's veracity has also been explored through the lens of natural, introspective language statements used in memory reporting. Justification statements, based on Tulving's (11) theoretical remember/know distinction,

Author affiliations: [a]Cognitive Science Program, University of Arizona, Tucson, AZ 85721; [b]Department of Psychology, University of Arizona, Tucson, AZ 85721; [c]Department of Pediatrics, School of Medicine, University of California, Irvine, CA 92697; and [d]Pulmonology Department, Children's Hospital of Orange County (CHOC), Orange, CA 92868

Author contributions: L.N. and K.C.S. designed research; K.C.S. performed research; K.C.S. analyzed data; and L.N. and K.C.S. wrote the paper.

Reviewers: H.L.R., Washington University in St. Louis; and J.T.W., University of California San Diego.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: nadel@arizona.edu or knsimon@uci.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2416373121/-/DCSupplemental.

Published December 19, 2024.

provide one avenue to determine a memory reporter's confidence in recalling their own experience or knowledge (11). Using machine learning, Dobbins & Kantner developed the Hits versus False Alarms (HFA) classifier to successfully use these justification statements to determine statement accuracy by separating words aligned to levels of confidence in recollection, such as I remember, from those referencing familiarity or guessing (12). Pursuing this further, Gamoran et al. (13) attempted to determine whether the human's memory veracity detection paralleled justification language classifiers by having participants rate the accuracy of a dataset that included participants' recorded memory responses during a word recognition test. This original test involved a word recognition, and for a subset of words that were correctly remembered, forgotten, correctly identified as new, or incorrectly identified as old, a required justification statement for the response. These justification statements were used as the stimuli in Gamoran et al.'s study where they found that their participants similarly used them to interpret the accuracy of others' memory. Further, they found individual differences in participants' ability to mentalize influenced their confidence in their own rating of the report. They theorized that the ability to mentalize may have led the reader to join the reporter, reinstating the context of the task and test. Thus, from these studies, we can glean that specific language cues can influence the veracity judgment of a reported memory.

Given that memories are not precise reproductions of experienced events and that fictional stories can be recounted as if they were true, determining the credibility of a reported memory is a crucial skill. Here, we introduce a unique approach to determining which elements of a reported memory impact its perceived credibility. In this initial exploration of the role of various kinds of details in influencing credibility, we used the Autobiographical Memory Interview rating system to create a series of memory narratives that varied in number and ratio of detail types. Our memory narratives varied in terms of how much of the content related to internal details that supported reinstatement of the experience, including context, place, and time, compared to those tangentially related to the experience. Our main hypothesis was that internal details, those directly relevant to the episodic nature of the study and memory being reported, would increase the credibility of the narrative as compared to external details and that the number would influence this rating. External details, those not directly related to the memory being reported, could tangentially be related without being experienced in the actual episode. We then followed up in a second study to investigate the impact of the quality of details, separating internal details into those directly referring to feelings, thoughts, or actions/reactions within the narrative and those implying context, time, space, or environment features.

Results

Study 1. Rating Analyses. We first investigated whether the type of detail, internal or external, influenced participants' ratings of the narratives using repeated measure ANCOVAs, with Greenhouse–Geisser correction reported when Mauchly's test indicated a violation in sphericity. For each individual rating, we evaluated within-subject factors of Narrative Type (Internal Only, External Only, Greater Internal, Greater External, and Equal Levels, Fig. 1 *A*–*E*). We covaried sex and BDI score for each analysis. *SI Appendix*, *Supplemental Materials* for analysis details for each individual rating.

Overall, we found a significant effect of Narrative Type for each individual rating [Accuracy: $F(1.517, 802.43) = 56.478$, $P < 0.001$, $\eta p^2 = 0.096$; Believability: $F(1.5, 793.75) = 70.68$, $P < 0.001$, $\eta p^2 = 0.118$; Confidence in the Truth: $F(1.52, 804.38) = 55.02$, $P < 0.001$, $\eta p^2 = 0.094$; Interesting: $F(2.276, 1204.09) = 164.178$, $P < 0.001$, $\eta p^2 = 0.237$; Pleasantness: $F(2.568, 1358.42) = 168.56$, $P < 0.001$, $\eta p^2 = 0.242$]. Further post hoc analyses revealed that

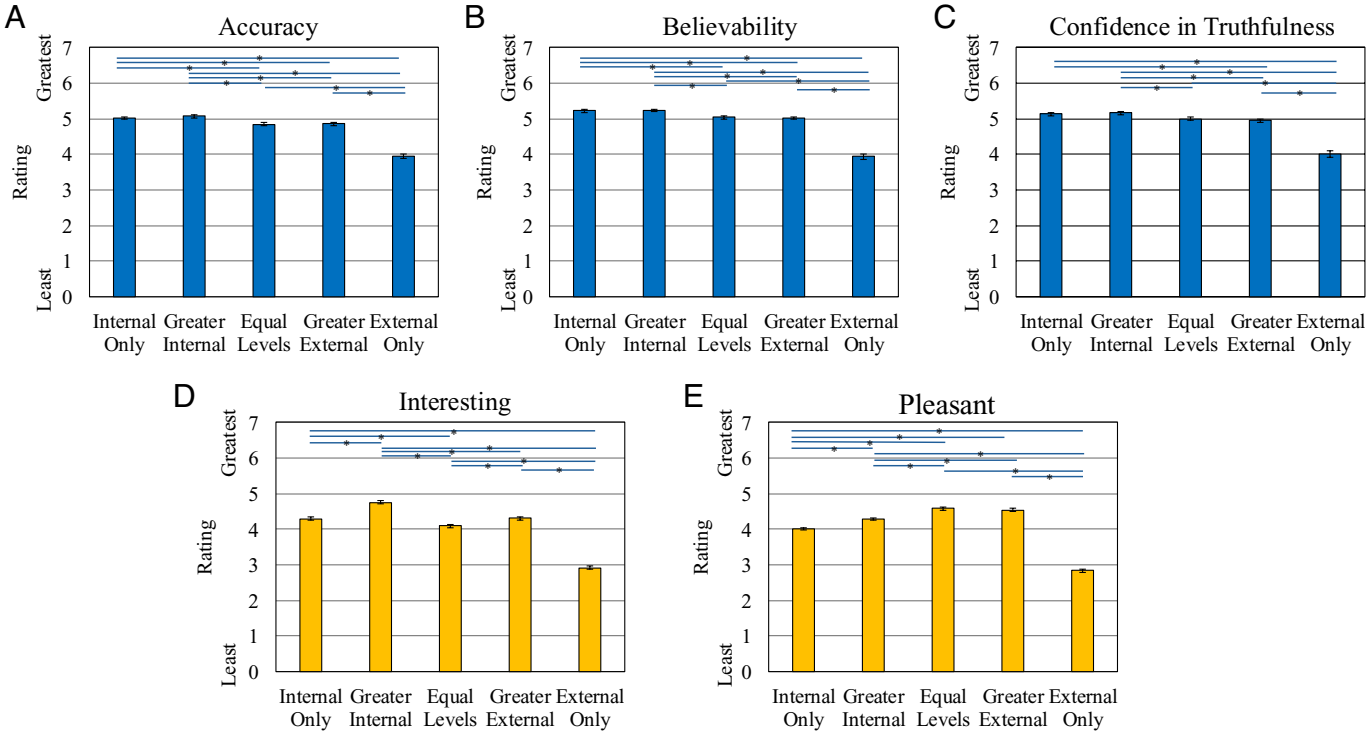

Fig. 1. Ratings for each narrative type. In the *Top* panel, individual ratings that align with the credibility composite score. From left to right are Accuracy (*A*), Believable (*B*), and Confidence in the Truth (*C*). In the *Bottom* panel, individual rating scales that align with saliency are Interesting (*D*) and Pleasant (*E*). Across all ratings, External Only was rating significantly lower than all other narrative types. Similarly, for credibility ratings, Internal Only and Greater Internal were rated higher than Greater External.

for each rating, External Only details were consistently rated significantly lower than all others.

Further, these initial individual rating analyses appear to show that narratives including internal details are rated more believable and accurate, while raters are more confident they are truthful. However, when rating narratives based on how pleasant or interesting they were, the external details became more influential, but only when they were mixed with some internal details. Narratives containing only external details were rated as least pleasant or interesting.

Composite Score Analyses. Given the thematic overlap of Accuracy, Believability and Confidence in the Truth as measures of overall credibility and the thematic overlap of Interesting and Pleasant as markers of saliency, we chose to create two composite scores. Prior to creating composite scores, we evaluated the internal consistency of credibility ratings across the narratives (*SI Appendix, Supplemental Materials*).

We investigated whether the number of internal or external details influenced participants' ratings of the narratives using a rmANCOVA with factors of Composite (Believability and Salience) and Narrative Type (Internal Only, External Only, Greater Internal, Greater External, and Equal Levels; controlling for sex and depressive symptoms, Fig. 2*A*).

We found significant main effects of Composite [$F(1, 529) = 194.42$, $P < 0.001$, $\eta p^2 = 0.269$] and Narrative Type [$F(1.53, 810.02) = 131.107$, $P < 0.001$, $\eta p^2 = 0.199$] and an interaction between the Composite and Narrative Type [$F(2.09, 1105.89) = 17.739$, $P < 0.001$, $\eta p^2 = 0.032$]. Credibility composite scores were rated significantly higher than salience composite scores ($P < 0.001$). We then carried out post hoc analyses to reveal the relationships between the detail levels (Fig. 2*B* for overview of narrative type relationships). For all narrative types, Credibility composite scores were rated significantly higher than Saliency composite scores (P's < 0.001). Within each composite, the pattern of scores paralleled the individual rating scales. For the credibility composite, Greater Internal narrative type was always rated significantly higher than other narrative types (P's < 0.001) except Internal Only ($P = 0.2$), External Only was always rated significantly lower than all other types (P's < 0.001), Internal-Only was also rated significantly higher than Equal Levels and Greater External (P's < 0.001), and there was no significant difference between External Only and Equal level ratings ($P = 0.468$). For the Saliency composite, we observed that Internal Only was rated significantly lower than all other narrative types (P's < 0.001) except greater than External Only ($P < 0.001$). External Only was rated significantly lower than all narrative types (P's < 0.001). Greater External was rated significantly higher than all other narrative types except Greater Internal (P's < 0.001). Lastly, Equal Level was rated significantly higher than Internal Only ($P < 0.001$). The analyses reported so far show that internal details, whether present alone or mixed with some external details, increased ratings for both credibility and salience.

Detail Number Analyses. We then investigated within each composite whether the number of internal and external details influenced rating scales separately for credibility and saliency (Fig. 3). We conducted rmANCOVAs for Narrative Type (Internal Only, External Only, Greater Internal, Greater External, and Equal Levels) and Detail Level (Low, Medium, or High). For believability, we found a significant interaction of Narrative Type, as reported above, a significant main effect for Detail Level [$F(2,1058) = 25.075$, $P = <0.001$, $\eta p^2 = 0.045$] and significant interaction between Narrative Type and Detail Level [$F(5.63, 2981.217) = 42.209$, $P < 0.001$, $\eta p^2 = 0.074$]. The post hoc analyses reported are Bonferroni corrected. For Internal Only narratives, the detail level ratings significantly differed from one another (P's < 0.001), with the highest ratings given to the narratives with fewest number of details, followed by those with the greatest number of details, and then by those with the middle number of details. There were no significant differences between External Only or Greater Internal detail levels (P's $= 1.0$). For Greater External narratives, the medium level of details was rated significantly lower than for narratives with the fewest and greatest number of details (P's < 0.001), but the fewest and greatest did not differ ($P = 1.0$). For the Equal Level narratives, those with the medium number of details were rated significantly higher than those with the lowest and highest number of details (P's < 0.011), and those with the greatest number of details were rated significantly higher than those with the fewest number of details ($P < 0.001$).

For salience, we found significant main effects of Narrative Type, as reported above, and Detail Level [$F(1.954, 1033.76) = 22.835$, $P = <0.001$, $\eta p^2 = 0.041$]. We also found a significant interaction between Narrative Type and Detail Level [$F(5.217, 2759.9) = 22.822$, $P < 0.001$, $\eta p^2 = 0.041$], and a significant interaction between Narrative Type, Detail Level, and Sex [$F(5.217, 2759.9) = 4.164$, $P < 0.001$, $\eta p^2 = 0.008$]. We also found a significant between-subjects effect of BDI [$F(1,529) = 4.009$, $P = 0.049$, $\eta p^2 = 0.008$]. The post hoc analyses reported are Bonferroni corrected. For Internal Only, all were significantly different from each other (P's < 0.001). The middle level

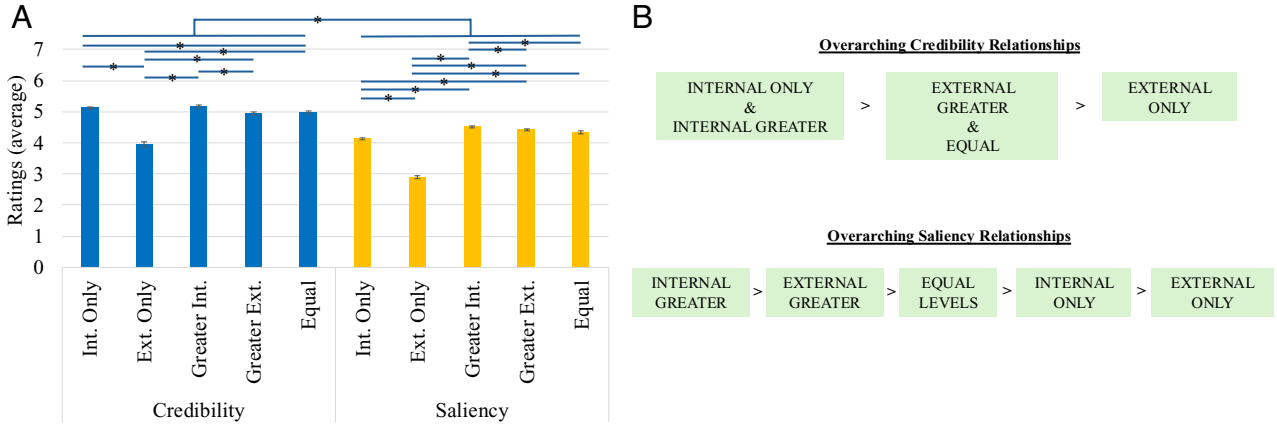

Fig. 2. Study 1 Internal v External Details. (*A*) Credibility and saliency composite scores are plotted for all narrative types. Overall, credibility composite scores were rated significantly higher than saliency scores. (*B*) Post hoc assessments of the relationships between narrative types are conceptually displayed in panel *B*. For both, Internal Only scores rated higher than External Only scores.

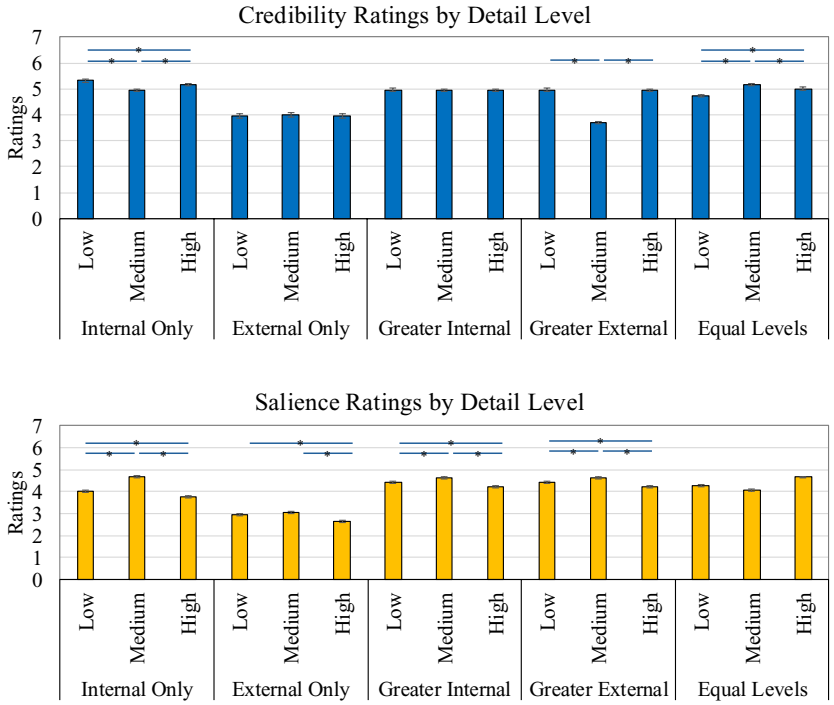

Fig. 3. Study 1 Composite scores evaluated by detail level. *Top* panel, credibility composite scores were compared for low, medium, and high detail numbers. *Bottom* panel, salience composite scores were compared across low, medium, and high detail levels. As can be seen, there is no linear relationship between the number of details and the score rating; however, differences exist across. Bonferroni corrections were applied for post hoc comparisons.

details had the highest saliency ratings, followed by the lowest level of details, and finally, the highest level of details. For External Only, the rating pattern remained the same numerically; however, there were only significant differences between the middle and highest detail level ratings and the highest and lowest detail ratings (*P*'s < 0.001). The difference between medium and low detail levels did not differ significantly ($P = 0.05$). For both Greater Internal and Greater External detail levels, medium level details were rated higher than low level, followed by highest level details (*P*'s < 0.001). For Equal Level details, the medium level of details were rated significantly lower than both low and high levels (*P*'s < 0.001) and lower level details were rated significantly lower than high level details ($P < 0.001$). Overall, these comparisons between high, medium, and low levels of both internal and external details show that both credibility and saliency were modulated by the number of details, though not in a linear fashion.

Study 2. Type of Internal Details. In our second study, we evaluated whether the types of internal details impacted our ratings. We compared two categories, Person-Related (PR) details that include explicit references to one's feelings or thoughts, or to other individuals present, and their actions/reactions within the narrative and Non-Person Related (NPR) details, which were those that referred to time, space, perceptual features of the environment, weather, or other events. We created a set of narratives that varied in the number of each of these kinds of details. Here, we present the analyses using composite scores for credibility and saliency for each PR and NPR internal narrative type, respectively. Prior to creating composite scores, we evaluated the internal consistency of credibility ratings across the narratives (*SI Appendix, Supplemental Materials*).

We ran a rmANCOVA with factors of Composite (Credibility and Salience) and Narrative Type (Non-Person-Related Only, Person-Related Only, Greater Non-Person-Related, Greater Person-Related; Equal Levels) with covariates of sex and depressive symptoms (Fig. 4). We found significant main effects of Composite [$F(1, 290) = 79.22$, $P < 0.001$, $\eta p^2 = 0.215$] with the credibility composite rated significantly higher than the salience composite. We also found a main effect of Narrative Type [$F(3.724, 1079.92) = 22.784$, $P < 0.001$, $\eta p^2 = 0.073$]. We also found an interaction of Composite and Narrative Type [$F(3.45, 1002.76) = 24.23$, $P < 0.001$, $\eta p^2 = 0.077$], demonstrated in Fig. 4. Finally, we found a significant three-way interaction of Composite by Narrative Type by depressive symptom, discussed further in *SI Appendix, Supplemental Materials*. All post hoc analyses were Bonferroni corrected. For the credibility composite, we found that Person-Related, Greater Person-Related, and Equal Level details were consistently rated as higher than other Non-Person-Related details (*P*'s < 0.001). And although the Non-Person-Related details were consistently rated significantly lower than Person-Related details, Non-Person-Related only and Greater Non-Person-Related details were not significantly different than each other ($P = 1.0$). Interestingly there was not the same clear-cut pattern for saliency composite scores. Person-Related only was significantly lower than all other details (*P*'s < 0.001) and Non-Person-Related only details were rated lower than all the other narrative types except Person-Related Only (*P*'s < 0.001). There was no significant difference between ratings for memories that had Greater Non-Person-Related details compared to Greater Person-Related details; however, equal levels for both narrative types rated the highest compared to all other types ($P < 0.001$).

We then compared credibility and saliency composite scores when only Non-Person-Related details or Person-Related details were present at low or high levels (see Fig. 5). For credibility composite, we found significant main effects of Composite [$F(1, 290) = 116.3$, $P < 0.001$, $\eta p^2 = 0.286$] and Detail Level [$F(1, 290) = 41.955$ $P < 0.001$, $\eta p^2 = 0.126$]. We did not find a main effect of Narrative Type ($P = 0.693$). We also found two significant interactions, between Composite and Narrative Type [$F(1, 290) = 43.6$, $P < 0.001$, $\eta p^2 = 0.131$] and between Composite and Detail Level [$F(1, 290) = 6.4$, $P = 0.012$, $\eta p^2 = 0.022$). We also found two three-way interactions, one between Composite, Detail Level, and Sex [$F(1,290) = 11.963$, $P < 0.001$, $\eta p^2 = 0.04$], and between Composite, Narrative Type, and Detail Level [$F(1,290) = 16.223$,

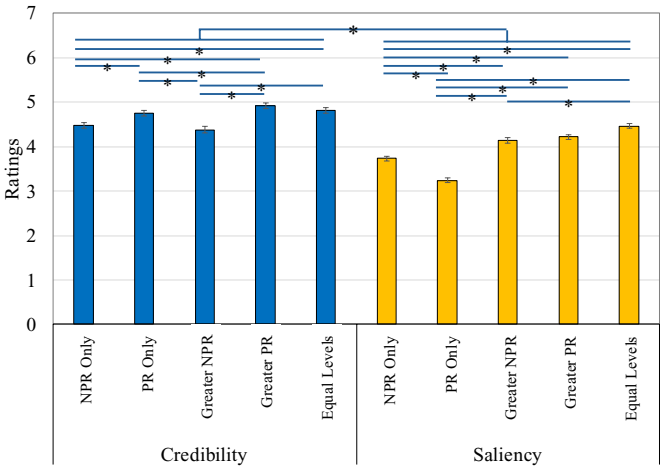

Fig. 4. Study 2. Internal Detail Stories. For credibility and salience composites, we compared two types of internal narrative types: Non-Person-Related internal details, consisting of time, space, perceptual features of the environment, weather, or external events, and Person-Related details consisting of internal feelings or thoughts, or to other individuals present, and their actions.

$P < 0.001$, $\eta p^2 = 0.053$]. We also found a significant between-subjects effect based on Sex [$F(1, 290) = 8.1941$, $P = 0.005$, $\eta p^2 = 0.027$]. Overall, credibility composite scores were rated higher than saliency composite scores and higher detail memories were rated higher than low detail memories. Within the credibility composite, Non-Person-Related details were rated significantly lower than Person-Related details ($P < 0.001$), while within the saliency composite, Non-Person-Related details were rated significantly higher than Person-Related details ($P < 0.001$). Within each composite high-level details were rated significantly higher than low-level details (P's < 0.001). For the credibility composite, at both the low and high detail levels, Non-Person-Related detail memories were rated significantly lower than Person-Related detail memories (P's < 0.009). However, for the salience composite, this relationship was only significant at the low ($P < 0.001$) but not high ($P = 0.097$) detail levels. As Fig. 5 shows, differences between the two types of internal details were consistently significantly different for Person-Related, but not Non-Person-Related details. Further, at this level of detail, the two types of internal details impacted credibility and saliency differently, a pattern we had not seen in our earlier data when we did not control the type of internal and external details used.

Discussion

Sharing memories is a functionally communicative act by humans with social and information propagating implications. Establishing the reliability and veracity of the information received is crucial to determining whether it can be useful in shaping future behavior. The two studies reported here explored how different types of episodic memory details might influence the likelihood that a memory narrative is believed. These studies sought to determine how internal and external details influenced subjective judgments of credibility. Our first study showed that internal episodic details directly tied to the narrative increased credibility ratings relative to external episodic details that were not directly relevant to the main narrative. In Study 2, we looked more carefully at different types of internal episodic details, distinguishing between details that referenced extraneous, Non-Person-Related perceptual features and details that were Person-Related, finding that the latter were most directly tied to higher credibility ratings.

Determining the credibility of information seems an essential task, but judging the veracity of others' statements is generally poor (14). Sources of information and the presentation of the information can influence believability (15–17). Behaviorally, there are few consistent behavioral cues to deception that judgers can rely on, and even these are only weakly related to deception (8, 18, 19). Moreover, these behavioral cues can manifest themselves in other emotional states and are not linked solely to deception. As such, determining the credibility of what one hears, or reads, in the absence of behavioral cues is important. In the forensic literature, a number of early reports showing that a greater degree of detail, termed vividness, enhanced believability assessments (20, 21). Bell & Loftus (22) further showed that the specificity of detail influenced believability. Their participants, acting as mock jurors, rated a series of prosecutor and eye witness court statements for believability and guilt. The statements varied in terms of the specificity and related or unrelated nature of included details. Detail specificity influenced believability judgments, whether the details were related or unrelated. The statement examples that Bell & Loftus provided in their articles suggest that both their related and unrelated details would be defined as internal details in our nomenclature, hence their having the same impact on believability makes sense. Other factors influencing the believability of witness statements include the consistency and logic of the reported memory and the confidence of the speaker (8, 23). It would be of interest to view these latter studies through the lens of internal vs. external details, with a focus on the specific kinds of internal details that seem to most powerfully influence believability.

Our first study demonstrated the overall impact of the presence of internal details on credibility judgments, but did not determine which types of internal details were most influential. When Levine et al. (10) introduced the dichotomy between Internal and External details to the memory literature, they included many different kinds of details in the Internal category. We used these definitions to create two groupings, based on whether the internal detail was primarily linked with the exterior or interior aspects of the narrative. We found that person-based details more strongly impacted credibility ratings. Person-Related details are almost always referable to single episodes, as they relate unique thoughts, feelings and reactions one had, or one observed in others during the episode being depicted. They are evanescent. Most Non-Person-Related details are different, referring to things that could be a part of many different events, as they are about features of the external world that are not evanescent. This is why schematic memories, that result from repeated episodes sharing significant commonalities, are generally rich in what we are calling external physical details. The fact that Person-Related details were rated as more believable than external physical details could suggest that people might react differently to memory narratives that are clearly about a single event, versus those that reflect what Neisser (24) called repisodic memory—memory for a kind of event we have experienced multiple times (24).

There has been some research comparing the believability of memories experienced once or multiple times (25, 26). Here, a set of participants experienced an event once or repeatedly and then retold that event (27). Another group of individuals rated these reported memories and appeared to rate memories for single events as more believable than those experienced repeatedly. This is what we would expect if the narratives produced by the participants who experienced an event multiple times had fewer Person-Related details. Moving forward, investigating how internal and external details influence the perception of reported traumatic events could be useful in understanding perceptions of an individual's credibility in judicial contexts. Future research can further evaluate how these detail ratios influence credibility perceptions within different clinical contexts. Our research was

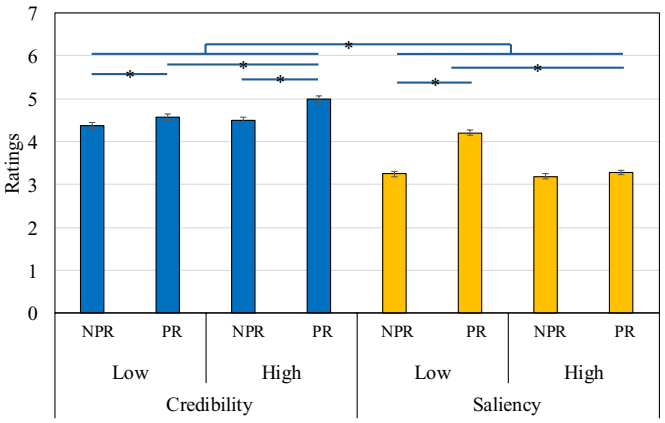

Fig. 5. Study 2. We evaluated how the number and type of internal details influenced ratings for credibility and salience composite scores. NPR refers to Non-Person-Related internal details, and PR refers to Person-Related internal details.

conducted during the earlier stages of the COVID-19 pandemic, when rates of psychiatric symptoms were known to be high, so any replication of our findings should take that into account.

Literature shows that recalling a memory may render it susceptible to errors, as details can be influenced by a host of factors including suggestibility, intervention, forgetting processes, and contextual reinstatement (2, 28–30). Experiments investigating the accuracy of true memories are methodologically difficult, relying on the reporter's perception of accuracy, real-life events, or second-party members to the event. Diamond et al. (31) also used Levine et al.'s Autobiographical Interviewing scoring system to experimentally assess recall accuracy of two distinct past events in younger and older adults. They found that both age groups reliably reported greater internal than external details of the event that they experienced. Further, factors such as age and time between event and recall negatively influenced the rate of recalled details. Of the internal details reported, they determined that the vast majority were accurate; however, the authors noted that for both groups, the number of details reported was significantly fewer than were independently coded as experienced by the researchers and that errors were reported by all. Thus, details of past real-world events could be recalled accurately, but the total number was low, and inaccurate details were also reported, just proportionally fewer. Interestingly, the authors then asked expert memory researchers their belief in the likelihood of accurate detailed reports of the memory, with the experts predicting high memory errors. This belief contradicts the data but is understandable as experimental conditions, such as forced reporting and misinformation, can increase errors. In our own study, memory narratives were fabricated to match specific ratios of internal and external details. Consequently, we were unable to investigate the link between accuracy and credibility based on narrative types. Future research could evaluate this link experimentally, in line with Diamond and colleagues' experimental design, expanding this detail rating into court testimony.

In court testimony, perceived credibility plays an important role in jurors' decision to convict yet is not always associated to the actual accuracy of the reported event. Factors such as the witness' age, sex, experience, emotionality, and confidence in reporting can influence the perception of credibility (see reviews 31, 32). The conflation of credibility with accuracy is observable in jurors' behavior, where confident testimony or identification of perpetrators can influence convictions (33). Often research finds errors in reporting, despite confidence in identification; however, recent work has found that confidence is related to accuracy on an initial lineup test (34) and that initial confidence is particularly informative when reaction time is also considered (33, 34). In our own study, memory narratives were fabricated to match specific ratios of internal and external details. Consequently, we were unable to investigate the link between accuracy and credibility based on narrative types. We recently became aware of a preprint describing an exploratory study showing that memory fidelity judgments were influenced by the specificity of memory details (35). As in our study, the more specific the details, the higher were the fidelity rankings. Future research could further evaluate this link experimentally, in line with Diamond and colleagues' experimental design, expanding detail rating into court testimony.

Recently, Dodson & Dobolyi (36, 37) investigated how confidence ratings are inferred from verbal utterances, specifically in the context of lineup decisions in the judicial system (35, 36). Their study design examined how people inferred an eyewitness' confidence in identification based on brief additional statements. Mock eyewitnesses read verbal expressions of confidence, some of which included a feature-based justification while others involved unobservable justification. Interestingly, perceived confidence was reduced when feature justification utterances were included, but not when the statements simply reported confidence or also included an unobservable justification. How these findings relate to our current study is intriguing and warrants further investigation. In our studies, a feature justification such as "I remember his chin" would be classified as a person-related internal detail. How the inference of confidence in a lineup decision with justification relates to credibility of the witness is unknown. Further, the degree to which brief, single line internal utterances impact credibility is unknown. Expanding upon Dodson and colleagues' study, it is possible that if reviewers were presented with the verbal expression within a longer memory report, the featural justification might not diminish the inferred confidence rating as much. Determining whether the rate or proportion of these utterances influence inferred confidence or credibility is an important avenue to explore further.

Perceived credibility is a nuanced phenomenon, shaped by numerous individual and contextual factors (32, 38, 39). As noted earlier, Dodson & Dobolyi discovered that the inclusion of feature-based justification statements reduced inferred confidence, even when reports of confidence were high (35). Perceptions of credibility may also differ across domains. In judicial contexts, the consequences of credibility perceptions are more significant, directly impacting decisions. Whether the same detail composition affects perceived credibility across settings is less understood. For instance, does the detail composition affect credibility when additional context is provided? Future studies should explore the role of context while manipulating detail composition. One potential next step could be to add characteristics about the memory reporter, such as age, sex, or race, or include a spectrum of familiarity, from close friend to stranger.

Our results might also have wider societal implications, in particular regarding discussions of misinformation in society. In a fascinating recent report concerning how to bridge political divides, Kubin et al. showed that "furnishing perceptions of truth within moral disagreements is better accomplished by sharing subjective experiences, not by providing facts" (40). While the studies leading to this conclusion did not look at the same kinds of listener reactions we investigated here, the leading role played by subjective experiences compared to generic facts seems quite similar to the leading role played by person-related details in our work. Further study of the interactions between the stories we tell each other and how we might best talk to, rather than at, those across a political divide are clearly warranted.

In a real-world political example, at the Congressional Watergate Hearings in the 1970s, John Dean provided vivid testimony about critical presidential conversations that implicated President Nixon in obstruction of justice. Some years later the accuracy of Dean's memories was assessed by comparing Dean's testimony to the subsequently released Watergate tape transcripts. Neisser (24) showed that though Dean was right about the gist of the conversations, he was wrong about many of the specific details. At the time, this was not known by the Congressional committee members, who instead were duly impressed by, and even commented on, Dean's command of the details of the events he was describing. Despite being relatively unknown by his observers at the time of his testimony, his extensive provision of details seems to have played a critical role in convincing his audience that his memories were true, in this case affecting both the perceived believability of the event and the credibility of an unfamiliar storyteller. How the detail composition of prior memories relates differentially to accuracy compared to judgments of credibility across different contexts will be an important step for future studies to explore. Further, at about the same time (1982), Ronald Reagan became the first President to include references to specific individuals and events, as part of his State of the Union address. Such inclusions are now habitual, and one cannot help but think that appealing to individuals and their stories does more than just engage one's audience emotionally. It might, given the present results, also lead the audience to find the speaker more credible.

Conclusion

There has been considerable research demonstrating how we form, store, and retrieve episodic memories; however, less attention has been given to how others perceive our retrieved memories. When we recount memories, the details we share can vary, some being quite specific and integral to the memory while others are more generic and tangential. Our studies revealed that memories with specific details were perceived as most believable, particularly those that involved person-related details. Collectively, our findings indicate that how we relate our past experiences can shape the judgments of others.

Materials and Methods

Participants. A total of 825 subjects participated in two, separate within-subject design studies are reported here. Study 1 had a final sample of 532 participants (349 female) with a mean age of 20.2 years (SD = 2.16)–72 additional participants did not complete this study and were discarded from all analyses. Study 2 had a sample of 293 participants (214 female) with a mean age of 19.04 (SD = 1.853). *SI Appendix*, Table S2 for each study's participant demographic information. These studies were conducted remotely using Qualtrics during COVID-19 and as one of the few available avenues for research participation, we did not exclude students nor restrict our participant number. Participants were recruited from the SONA system attached to psychology courses at the University of Arizona and received course credits for their time. Prior to participating, all participants provided written consent and were then provided a Qualtrics link to complete the study. Our studies and procedures were approved by the Institutional Review Boards at the University of Arizona and the University of California, Irvine.

Study Design. Participants in both studies followed the same procedure: Each read a series of narratives and rated each one on five distinct Likert scales for accuracy, believability, interest, pleasantness, and confidence in the truthfulness of the narrative (Fig. 6 for study design depiction). Narrative order was randomized across participants. Each narrative appeared in isolation with the following questions posed below: How interesting is this memory? How accurate do you believe this memory is? How believable is this memory? How pleasant is this memory? How confident are you that this memory is truthful?

For each, participants completed a Likert scale from 1 to 7, with 1 being not at all, 4 being somewhat, and 7 being very. The aim of these questions was to determine the readers' credibility assessment using the responses to accuracy, believability, and confidence in the truth, and separately, their saliency assessment using the responses to questions about interestingness and pleasantness. The narratives varied with regard to the number and types of internal and external details, as described below (*SI Appendix*, Table S3 and *Supplemental Materials* for examples of the narrative stimuli for different narrative types). Participants were also administered a demographic screener and empirically validated questionnaires to assess depressive, anxiety, memory symptoms, and personality factors.

Study 1 Narrative Stimuli. In Study 1, each participant was administered 30 narratives that varied in terms of the number of internal and external details it contained. The narrative stimuli met the following conditions: composed of only Internal only (Internal Only), composed of only External details (External Only), composed of greater Internal than External details (Greater Internal) at ratios of 2 to 1, approximately 3 to 2, and 3 to 1, greater External than Internal details (Greater External) similarly at ratios of 1 to 2, approximately 2 to 3, and 1 to 3, or the same number of details, a one to one ratio, in each (Equal Levels).

Within each of these narratives, we further varied the number details between 0, 6, 12, or 18, which we called detail levels (Fig. 7 for graphical depiction). This allowed us to evaluate whether the total number of details influenced credibility ratings. For example, Internal Only narratives could contain a total of either 6, 12, or 18 episodic, internal story details. These details could include specific references to context, time, and place experienced in the memory being reported. For example, this sentence includes 4 internal details: "Last year, when I attended Brew at the Zoo, I fed a giraffe some leaves while drinking an IPA." This sentence links to a specific time, place, and two activities the reporter engaged in at the time. Internal Only narratives never included a single tangential, external detail. In parallel, External Only narratives could contain a total of 6, 12, or 18 external details that were extraneous and tangentially related to a narrative topic but did not include specific references to an experience that could be reinstated. For example, this sentence includes three external details: "My grandpa was a pilot, which is how he figured out that he that liked America." This sentence is included within a memory report of traveling. This is not directly experienced but does note a person, thought, and location. Equal number narratives had the same number one-to-one ratio of internal and external details, thus could have 6 internal and 6 external details intermixed, 12 of each intermixed, or 18 of each intermixed. Greater Internal detail narratives had both internal and external details, with internal details always being greater, i.e., 18 internal details to either only 6 or 12 external details, or 12 to 6 details, (i.e., 3-to-1, 3-to-2, or 2-to-1 ratios).

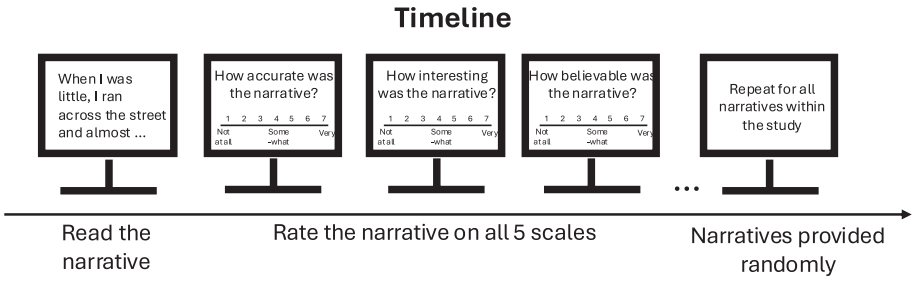

Fig. 6. Schematic of timeline. All participants read 30 narratives and rated each narrative on 5 scales: accuracy, believability, confidence in the truth, interest, and pleasantness. After, participants completed a demographic screener and questionnaires on health and mood.

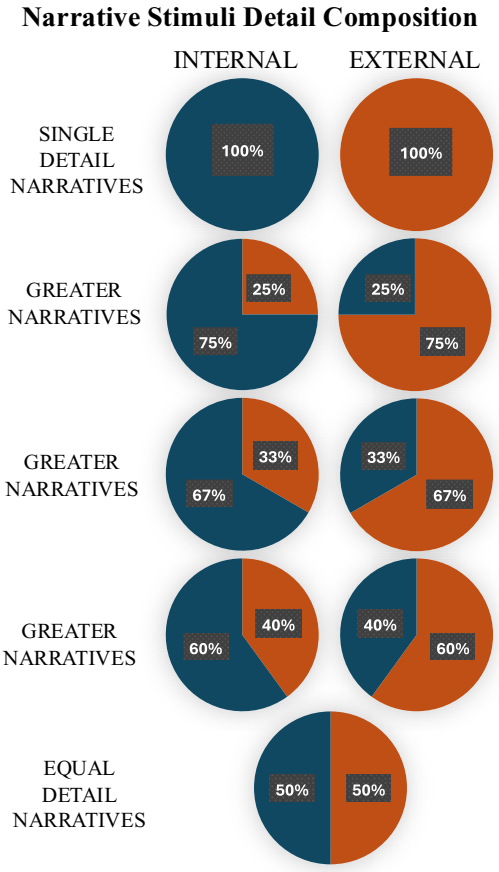

Fig. 7. Study 1 memory narrative types graphically presented. Narratives were composed of internal and external details that differed in detail ratios. In the single detail narratives, memories were composed of only Internal or only External details. The following three types of narrative were composed of differing internal/external detail ratios, included a 1:3 ratio, an approximate 1:2 ratio, and a 2:3 ratio. Greater Internal detail narratives always had more internal than external details, and Greater External detail narratives always had more external than internal details. Our last narrative stimuli had equal quantities of internal and external details intermixed. Participants each read and rated the set of narratives which included at least two representations of all narrative types.

Greater External detail narratives had the opposite ratios. For each detail level, we created two narratives, which were averaged to be a single score. Each participant was administered the exact same set of narratives, in an order randomized across participants, within a single experimental session.

Word count was kept consistent across narratives of each detail level, to the extent possible. For example, the two Internal Only 6 detail level narratives had 57 and 61 internal details and the two External Only 6 detail level narratives had 58 and 62 words (M = 59.5 words, SD = 2.3). Across all Study 1 narratives, the word count ranged from 58 to 141 words. *SI Appendix*, Table S3 for all narrative word counts.

Study 2 Narrative Stimuli. In Study 2, participants were administered a set of 10 Internal Only narratives, which varied types of internal details. Here, we distinguished, and separately varied, two types of Internal details: PR details which included references to ones' own feelings/thoughts, or to individuals present and their behaviors or presumed feelings/thoughts and NPR which included references to context, time, space, the weather, perceptual features, or concomitant, but Non-Person-Related events. These narratives included 0, 3, or 6 details of each type (PR or NPR). Two narratives were created for each narrative type and were averaged together to form a single score. Thus, the narratives were composed of either only NPR details, only PR details, a ratio of 1-to-2 greater NPR details to PR details, or the inverse with greater PR to NPR details, or equal numbers of each.

The word count for Study 2 narratives was also kept consistent (*SI Appendix*, Table S3); however, as fewer details were included in the stories, the count ranged from 28 details to 102 details. As in Study 1, we kept narratives of the same detail level consistent within this range. For example, for the two narratives of 6 NPR details and 3 PR details, the counts were 68 and 73, while the opposite narratives, 3 NPR details and 6 PR details were a total of 68 and 74 words.

Questionnaires. Participants completed a demographic survey reporting sex, race/ethnicity, age, education, self-identified reading level, household income, and psychological history (*SI Appendix*, Table S2 for demographics). Participants also completed empirically validated questionnaires to determine mood symptoms, memory concerns, and personality factors. Participants completed: the Beck Depression Inventory-II, a 21-item questionnaire in which participants self-reported symptoms of depression on a Likert scale from 0 to 3, with higher numbers corresponding to more severe depressive symptoms (41); the Generalized Anxiety Disorder (GAD-7), a 7-item questionnaire in which participants self-reported symptoms of anxiety on a Likert scale with higher numbers corresponding to more severe anxiety (42); The Everyday Memory Questionnaire-Revised (EMQ-R), a 28-item questionnaire to assess subjective every day memory failures (43). Finally, participants also completed the Ten Item Personality Inventory (TIPI) to assess degrees of extraversion, agreeableness, conscientiousness, emotional stability, and openness (44). *SI Appendix*, Table S2 for questionnaire data.

Statistical Analyses. For Study 1, we first analyzed participants' ratings of the narratives on each accuracy, believability, interest, pleasantness, and confidence in the truth scales. As each detail level had a pair of narratives (i.e., two Internal Only 6 detail narratives), we averaged participants' ratings of the two (i.e., both scale accuracy ratings were averaged for Internal Only narratives with 6 details). Then, for each individual rating scale, we averaged all the detail levels within a narrative type together. In the example of the Internal Only narratives, we averaged the detail levels of 6, 12, and 18 details such that we had a single score per person per rating. This allowed us to control for the word count differences across levels which were equivalent across narrative types. We then ran a repeated measure analysis of covariance (rmANCOVA) for each scale type using within-subject factors of Narrative type (Internal Only, External Only, Greater Internal, Greater External, and Equal Levels). For all analyses, we covaried sex and BDI score as data collection coincided with the COVID-19 pandemic, an identified time of increased depressive symptoms. We corrected using the Greenhouse–Geisser method when Mauchly's test indicated a violation in sphericity.

We then grouped our scales into two composite scores: a credibility score, which included the accuracy, believability, and confidence in the truth ratings; and a saliency score, which included the interest and pleasantness ratings. To calculate these composites, we used the average narrative type from the first analysis and averaged the ratings for each composite. For example, for the credibility composite we averaged the rating groupings of accuracy, believability, and confidence in truth and for the saliency composite we averaged the rating groupings for interest and pleasantness. We used SPSS version 29 for all analyses. We ran repeated measure analysis of covariance (rmANCOVA) using within factors of Composite (Credibility, Salience) and Detail Level (names listed above) and covaried sex and depressive symptoms. For all significant findings, we ran post hoc analyses and applied Bonferroni corrections. When sphericity was significant, we corrected using the Greenhouse–Geisser method. We also evaluated whether the detail level (i.e., number of details) influenced the composite score by conducting an rmANCOVA of Composite (Credibility and Saliency) by Narrative Type (Internal Only, External Only, Greater Internal, Greater External, and Equal Levels) by Detail Level (Low, Medium, or High). We also ran reliability analyses for the individual ratings within a narrative type to support the formation of overarching composites for credibility and saliency. These composites provided us the opportunity to evaluate the overarching patterns across the measures, with greater reliability.

Finally, we correlated the relationships between psychiatric symptom levels (depressive symptoms, anxiety symptoms, memory complaints) and personality factors with credibility and salience composite scores. Bonferroni corrections were applied within each composite score. *SI Appendix, Supplemental Materials* for these analyses.

In Study 2, we investigated the differences in ratings for narratives comprising two types of internal details. We used narratives that had Only NPR, Only PR or a ratio of these two types. We followed the same analytic plan by averaging the two narratives at each detail level on each of the individual rating scales of accuracy, believability, interest, pleasantness, and confidence in the truth.

Then, for each narrative type, we averaged all the detail levels for each rating scale, respectively. We first ran rmANCOVAs using within factors of Narrative Type (NPR details only, PR details only, Greater NPR details, Greater PR details, Equal Levels) and covariates of sex and BDI score. Then, we similarly combined thematic scales and conducted rmANCOVAs with Composite (Credibility and Salience) and Narrative Type (Non-Person-Related, Person-Related details) by Detail Level (Low or High).

Data, Materials, and Software Availability. All data is available at the OSF repository: https://osf.io/q574e/?view_only=c91f5ce3eb514f6d86c2a03753164229 (45). All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Katarina Krstic who supported pilot data collection while an undergrad at the University of Arizona. This work was supported by University of Arizona and K08 HD107161 from NIH-NICHD to K.C.S.

1. N. Burgess, E. A. Maguire, J. O'Keefe, The human hippocampus and spatial and episodic memory. *Neuron* **35**, 625–641 (2002).
2. L. Nadel, A. Hupbach, R. Gomez, K. Newman-Smith, Memory formation, consolidation and transformation. *Neurosci. Biobehav. Rev.* **36**, 1640–1645 (2012).
3. N. Alea, S. Bluck, Why are you telling me that? A conceptual model of the social function of autobiographical memory *Memory* **11**, 165–178 (2003).
4. S. B. Klein *et al.*, Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Soc. Cogn.* **27**, 283–319 (2009).
5. J. B. Mahr, G. Csibra, Why do we remember? The communicative function of episodic memory *Behav. Brain Sci.* **41**, e1 (2018).
6. M. K. Johnson, Memory and reality. *Am. Psychol.* **61**, 760 (2006).
7. M. K. Johnson, J. G. Bush, K. J. Mitchell, Interpersonal reality monitoring: Judging the sources of other people's memories. *Soc. Cogn.* **16**, 199–224 (1998).
8. S. L. Sporer, S. J. Sharman, Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events *Appl. Cogn. Psychol. Off. J. Soc. Appl. Res. Mem. Cogn.* **20**, 837–854 (2006).
9. D. B. Wright, A. Memon, E. M. Skagerberg, F. Gabbert, When eyewitnesses talk. *Curr. Dir. Psychol. Sci.* **18**, 174–178 (2009).
10. B. Levine, E. Svoboda, J. F. Hay, G. Winocur, M. Moscovitch, Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychol. Aging* **17**, 677 (2002).
11. E. Tulving, Remembering and knowing the past. *Am. Sci.* **77**, 361–367 (1989).
12. I. G. Dobbins, J. Kantner, The language of accurate recognition memory. *Cognition* **192**, 103988 (2019).
13. A. Gamoran, L. Lieberman, M. Gilead, I. G. Dobbins, T. Sadeh, Detecting recollection: Human evaluators can successfully assess the veracity of others' memories. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2310979121 (2024).
14. C. F. Bond, B. M. DePaulo, Accuracy of deception judgments. *Personal. Soc. Psychol. Rev.* **10**, 214–234 (2006).
15. J. K. Burgoon *et al.*, Interactivity in human–computer interaction: A study of credibility, understanding, and influence. *Comput. Hum. Behav.* **16**, 553–574 (2000).
16. M. D. Slater, D. Rouner, How message evaluation and source attributes may influence credibility assessment and belief change. *Journal. Mass Commun. Q.* **73**, 974–991 (1996).
17. C. N. Wathen, J. Burkell, Believe it or not: Factors influencing credibility on the Web. *J. Am. Soc. Inf. Sci. Technol.* **53**, 134–144 (2002).
18. B. M. DePaulo *et al.*, Cues to deception. *Psychol. Bull.* **129**, 74 (2003).
19. M. Hartwig, C. F. Bond Jr., Lie detection from multiple cues: A meta-analysis. *Appl. Cogn. Psychol.* **28**, 661–676 (2014).
20. R. M. Reyes, W. C. Thompson, G. H. Bower, Judgmental biases resulting from differing availabilities of arguments. *J. Pers. Soc. Psychol.* **39**, 2 (1980).
21. J. Shedler, M. Manis, Can the availability heuristic explain vividness effects? *J. Pers. Soc. Psychol.* **51**, 26 (1986).
22. B. E. Bell, E. F. Loftus, Trivial persuasion in the courtroom: The power of (a few) minor details. *J. Pers. Soc. Psychol.* **56**, 669 (1989).
23. R. P. Fisher, N. Brewer, G. Mitchell, "The relation between consistency and accuracy of eyewitness testimony: Legal versus cognitive explanations" in *Handb. Psychol. Investig. Interviewing Curr. Dev. Future Dir.* R. Bull, T. Val, T. Williamson Eds (2009). pp. 121–136.
24. U. Neisser, John Dean's memory: A case study. *Cognition* **9**, 1–22 (1981).
25. D. A. Connolly, H. L. Price, J. A. A. Lavoie, H. M. Gordon, Perceptions and predictors of children's credibility of a unique event and an instance of a repeated event. *Law Hum. Behav.* **32**, 92–112 (2008).
26. T. P. M. Theunissen, T. Meyer, A. Memon, C. C. Weinsheimer, Adult eyewitness memory for single versus repeated traumatic events. *Appl. Cogn. Psychol.* **31**, 164–174 (2017).
27. C. C. Weinsheimer, P. I. Coburn, K. Chong, C. L. MacLean, D. A. Connolly, Perceptions of credibility for a memory report of a single versus repeated event. *Appl. Cogn. Psychol.* **31**, 414–423 (2017).
28. E. F. Loftus, H. G. Hoffman, Misinformation and memory: The creation of new memories. *J. Exp. Psychol. Gen.* **118**, 100 (1989).
29. M. J. Sekeres *et al.*, Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learn. Mem.* **23**, 72–82 (2016).
30. H. L. Roediger, M. Abel, The double-edged sword of memory retrieval. *Nat. Rev. Psychol.* **1**, 708–720 (2022).
31. N. B. Diamond, M. J. Armson, B. Levine, The truth is out there: Accuracy in recall of verifiable real-world events. *Psychol. Sci.* **31**, 1544–1556 (2020).
32. B. A. Spellman, E. R. Tenney, Credible testimony in and out of court. *Psychon. Bull. Rev.* **17**, 168–173 (2010).
33. J. T. Wixted, L. Mickes, R. P. Fisher, Rethinking the reliability of eyewitness memory. *Perspect. Psychol. Sci.* **13**, 324–335 (2018).
34. A. Quigley-McBride, G. L. Wells, Eyewitness confidence and decision time reflect identification accuracy in actual police lineups. *Law Hum. Behav.* **47**, 333 (2023).
35. C. Bastin, A. Folville, M. Geurten, "I trust you if your memory is detailed": Interpersonal memory fidelity judgments and social bonding. Available SSRN 4303218 (2022).
36. C. S. Dodson, D. G. Dobolyi, Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law Hum. Behav.* **39**, 266 (2015).
37. C. S. Dodson, D. G. Dobolyi, Judging guilt and accuracy: Highly confident eyewitnesses are discounted when they provide featural justifications. *Psychol. Crime Law* **23**, 487–508 (2017).
38. J. van Doorn, N. N. Koster, Emotional victims and the impact on credibility: A systematic review. *Aggress. Violent Behav.* **47**, 74–89 (2019).
39. G. S. Goodman, J. M. Golding, V. S. Helgeson, M. M. Haith, J. Michelli, When a child takes the stand: Jurors' perceptions of children's eyewitness testimony. *Law Hum. Behav.* **11**, 27–40 (1987).
40. E. Kubin, C. Puryear, C. Schein, K. Gray, Personal experiences bridge moral and political divides better than facts. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2008389118 (2021).
41. A. T. Beck, R. A. Steer, G. K. Brown, *Beck Depression Inventory (BDI-II)* (Pearson, 1996).
42. R. L. Spitzer, K. Kroenke, J. B. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).
43. J. Royle, N. B. Lincoln, The everyday memory questionnaire–revised: Development of a 13-item scale. *Disabil. Rehabil.* **30**, 114–121 (2008).
44. S. D. Gosling, P. J. Rentfrow, W. B. Swann Jr., A very brief measure of the big-five personality domains. *J. Res. Personal.* **37**, 504–528 (2003).
45. K. C. Simon, L. Nadel, Data from "Perceived Memory Credibility". Open Science Framework. https://osf.io/q574e/?view_only=c91f5ce3eb514f6d86c2a03753164229. Deposited 3 December 2024.