I remember therefore I am: Episodic memory retrieval and self-reported trait empathy judgments in young and older adults and individuals with medial temporal lobe excisions

Caspian Sawczak a,*, Mary Pat McAndrews a,b, Brendan Bo O'Connor c, Zoë Fowler c, Morris Moscovitch a,d

a Department of Psychology, University of Toronto, Canada
b Krembil Research Institute at Toronto Western Hospital, University Health Network, Canada
c Department of Psychology, University at Albany, State University of New York, USA
d Rotman Research Institute at Baycrest Hospital, Canada

ARTICLE INFO
Keywords:
Episodic memory
Empathy
Self-knowledge
Aging
Medial temporal lobes

ABSTRACT
How do we know what sort of people we are? Do we reflect on specific past instances of our own behaviour, or do we just have a general idea? Previous work has emphasized the role of personal semantic memory (general autobiographical knowledge) in how we assess our own personality traits. Using a standardized trait empathy questionnaire, we show in four experiments that episodic autobiographical memory (memory for specific personal events) is associated with people's judgments of their own trait empathy. Specifically, neurologically healthy young adults rated themselves as more empathic on questionnaire items that cued episodic memories of events in which they behaved empathically. This effect, however, was diminished in people who are known to have poor episodic memory: older adults and individuals who have undergone unilateral excision of medial temporal lobe tissue (as treatment for epilepsy). Further, self-report ratings on individual questionnaire items were generally predicted by subjectively rated phenomenological qualities of the memories cued by those items, such as sensory detail, scene coherence, and overall vividness. We argue that episodic and semantic memory play different roles with respect to self-knowledge depending on life experience, the integrity of the medial temporal lobes, and whether one is assessing general abstract traits versus more concrete behaviours that embody these traits. Future research should examine different types of self-knowledge as well as personality traits other than empathy.

1. Introduction
Philosophers and psychologists alike have long recognized that there is a deep connection between what we remember and who we are (Klein, 2012; Prebble, Addis, & Tippett, 2013). Locke (1894), for example, went so far as to argue that a person is the same being that she was in the past only if she can recall past experiences. Other philosophers, such as Thomas Reid and Joseph Butler, proposed a more moderate view that the recollection of past experiences plays a central role in maintaining one’s sense of being the same person over time (Klein & Nichols, 2012). While philosophers have long recognized the importance of memory in self-knowledge, modern cognitive psychology has built a more fine-grained understanding that proposes distinct roles for different types of cognitive-neural memory systems in forming a sense of who we are. Broadly, psychologists refer to the recollection of specific personal experiences as “episodic memory” which is supported by a network of regions strongly connected to the medial temporal lobes (MTL) and the hippocampus in particular (Tulving, 2002), and it is distinguished from “semantic memory”, or memory for facts and general knowledge abstracted from the spatiotemporal context in which it was acquired, which is supported by a network of regions strongly connected to the lateral temporal cortex (Tulving, 1972). In this paper, we focus on how

Abbreviations: MTL, medial temporal lobes; mTLE, medial temporal lobe excisions; TEQ, Toronto Empathy Questionnaire; SMIS, Spontaneous Memory-Identity Survey.
* Corresponding author.
E-mail address: caspian.sawczak@alum.utoronto.ca (C. Sawczak).

https://doi.org/10.1016/j.cognition.2022.105124
Received 21 December 2020; Received in revised form 1 April 2022; Accepted 9 April 2022
Available online 25 April 2022
0010-0277/© 2022 Elsevier B.V. All rights reserved.
episodic memory may influence one’s sense of being an empathic person in young and older adults, as well as in people with MTL damage who have impaired episodic memory.

Particularly influential work in this area has been led by Martin Conway and colleagues (e.g., Conway, 2005; Conway, Justice, & D’Argembeau, 2019; Conway & Pleydell-Pearce, 2000; Conway & Rubin, 1993), arguing that the sense of self is primarily supported by semantic memory that has been abstracted from individual episodic memories and organized conceptually to form life narratives and trait summaries. Support for this theory comes from research demonstrating, for example, that the recollection of episodic memories does not facilitate performance (response latency) on a task in which people must decide whether a trait adjective accurately describes them (and vice versa), leading to the conclusion that episodic memory and self-knowledge are functionally independent (for reviews see Klein & Lax, 2010; Klein, Robertson, Gangi, & Loftus, 2008). Furthermore, damage to the hippocampus, which can cause pronounced deficits in episodic memory (Penfield & Milner, 1958; Squire, 1977) does not impair all forms of self-knowledge. For example, one study showed that the patient K.C., whose hippocampus was severely damaged after a head injury, rendering him completely amnesic in terms of episodic memory, was able to rate himself (and his mother) on various traits such that his ratings were consistent with those made by his mother (Tulving, 1993; see also Rosenbaum et al., 2005). Similar findings have been reported regarding the amnesic patient D.B. (Klein, Rozendal, & Cosmides, 2002) and in cases of developmental amnesia, where hippocampal damage has occurred neonatally (Picard et al., 2013). More recent work has found that both healthy individuals and patients with Alzheimer’s disease, a population with extensive brain damage including the hippocampus, show a similar increase in recalling semantic memories after being cued with self-images (e.g., “I am a mother”, “I am a student”), despite Alzheimer’s patients having severe episodic memory deficits and global cognitive decline compared to healthy controls. Such findings are taken as strong evidence that semantic memories, but not episodic memories, play a fundamental role in supporting self-knowledge (Rathbene et al., 2019).

Importantly, however, much of this large body of work is based on people’s ratings of how strongly (e.g., “to what extent”; Tulving, 1993) they identify with either single, abstract, trait adjectives (e.g., brave, popular) or self-images (e.g., “I am a mother”, “I am a student”). As pointed out by Prebble et al. (2013), it is conceivable that more concrete self-knowledge, such as how one might behave in particular situations, or perhaps how frequently one does so, is more reliant on episodic memory (see also Rathbone, Moulin, & Conway, 2008), but empirical work investigating this possibility has been lacking. A reliance on episodic memory may occur when retrieving self-knowledge invites consideration of highly contextual information about how one thinks, feels, and behaves in specific situations, which is the purview of episodic memory, as opposed to the more general fact-based semantic memory (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002; Moscovitch, Cabeza, Winocur, & Nadel, 2016; Repout, Davidson, Palombo, Moscovitch, & Levine, 2012; Repout, Irish, Moscovitch, & Rugg, 2019; Winocur, Moscovitch, & Bontempi, 2010). Intriguingly, many theories of personality and disposition in social psychology draw on widely used trait questionnaires comprised of items that appear to tap this more concrete, highly contextual type of self-knowledge (e.g., The Big Five; Costa Jr. & McCrae, 2008). Yet, the role of episodic memory in supporting this type of self-knowledge has largely been unexplored. There have been two case studies using versions of the Big Five personality questionnaire with individuals who had either developmental amnesia (Halliova, Addis, & Rosenbaum, 2020) or widespread acquired brain damage (Philippi et al., 2012). Both studies indicated similarity between patient and control samples along various dimensions (e.g., beliefs about oneself changing over time; reliability in self-report ratings). However, in the latter case, the patient’s family members noted that his ratings were consistent with his pre-injury personality but failed to reflect subsequent changes, which may indicate a diminished ability to update self-knowledge due to episodic memory impairments (Philippi et al., 2012; for a similar discussion of an individual with Alzheimer’s disease see Klein, Cosmides, & Costabile, 2003). Neither of these case studies directly explored associations of episodic memory retrieval and personality ratings.

There is some prior literature that suggests a direct relation between episodic memory retrieval and self-knowledge or sense of self. For example, one study found that writing a detailed description of a single, personally relevant episodic memory led people to generate a greater number of self-defining statements in an identity task compared with writing about a control topic (Charlesworth, Allen, Havelka, & Moulin, 2016). In another study, people reported using their autobiographical memories, especially those from adolescence and early adulthood, to maintain a sense of self-continuity (Wolf & Zimblish, 2016). In contrast to the Alzheimer’s findings reported above (Rathbone et al., 2019), Addis and Tippett (2004) found that people with Alzheimer’s produced fewer and vaguer statements about their identity when compared with age-matched healthy controls, and these differences were associated with their impoverished autobiographical memory, including the recollection of specific events—but not with global cognitive decline (see also Tippett, Prebble, & Addis, 2018). A study of people with MTL epilepsy, who are known to have deficits in episodic autobiographical memory (e.g., Penfield & Milner, 1958; St-Laurent, Moscovitch, Jadd, & McAndrews, 2014; St-Laurent, Moscovitch, Levine, & McAndrews, 2009; for review see McAndrews, 2012), showed that these individuals had an impoverished sense of personal identity (scoring lower on a scale of identity “exploration”) compared to both healthy controls and people with other forms of epilepsy (Allebone, Rayner, Siveges, & Wilson, 2015).

Older adults (aged 65+), as well, exhibit reductions in MTL structure and function (e.g., Berron et al., 2018; Jemigan et al., 2001; Persson et al., 2012; Raz et al., 2005), and these age-related changes are associated with deficits in episodic memory (for reviews see Hedden & Gabrieli, 2004; Park & Gutchess, 2005). Specifically, the episodic content of older adults’ autobiographical memories is lacking in episodic detail (e.g., Addis, Wong, & Schacter, 2008; Levine et al., 2002; Piolino et al., 2010) and in “autonoetic consciousness” (Boujut & Clarys, 2016; Piolino et al., 2006; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007), which is the subjective sense of mentally traveling back in time and re-experiencing a memory while recollecting it (Tulving, 1985). The semantic components of their autobiographical memories, however, are relatively preserved (e.g., Levine et al., 2002). Accordingly, the sense of self does not appear to be diminished in older adulthood, and in fact may be clearer and more stable (Rice & Pasupathi, 2010), consistent with the notion that older adults are no longer in the acute phase of identity development faced by young adults (Cramer, 2017; Erikson, 1956; Kroger, Martinussen, & Marcia, 2010). In addition, older adults’ sense of self seems to be less informed by individual episodic memories (Rice & Pasupathi, 2010; Wolf & Zimblish, 2015, 2016), perhaps because they have accumulated enough life experiences for statistical regularities and themes to be abstracted from them in the form of semantic memory even for specific highly contextual situations (Moscovitch et al., 2016, 2005; Umnath & Marsh, 2014). Interestingly, while one study did find a positive correlation between the specificity of older adults’ episodic autobiographical memory retrieval and the strength or clarity of their sense of personal identity (measured by questions such as, “I know what I look like”, “I know what my morals are”; “I know what I want from life”), the authors also found evidence that this relationship was mediated by the older adults’ semantic self-knowledge (Haslam, Jetten, Haslam, Pugliese, & Tonks, 2011).

To investigate whether episodic autobiographical memory plays a role in specific trait self-knowledge, the present study examined healthy young adults as well as healthy older adults and people with MTL damage as they made self-reported trait empathy ratings on individual items belonging to the Toronto Empathy Questionnaire (TEQ; C. Sawczak et al. Cognition 225 (2022) 105124
Spreng, McKinnon, Mar, & Levine, 2009). The TEQ involves rating how frequently one exhibits empathy-related behaviours (e.g., “I get a strong urge to help when I see someone who is upset”). We chose to focus on a measure of trait empathy because concern for others’ welfare has already been linked to episodic memory and the MTL in previous work (e.g., Beadle, Tranel, Cohen, & Duff, 2013; Ciaramelli, Bernardi, & Moscovitch, 2013; Gaesser, 2020). For example, encouraging participants to recall events in as much detail as possible, which is known to involve the MTL, leads them to express greater empathy to strangers in hypothetical scenarios (Vollberg, Gaesser, & Cicara, 2021). Further, when neurologically healthy people are asked to remember past instances of themselves helping others in need, they report being more willing to help in hypothetical situations (Gaesser, Hirschfeld-Kroen, Wasserman, Horn, & Young, 2019, Exp. 1; Gaesser & Schacter, 2014, Exp. 3), but when older adults and people with MTL excisions engage in a similar activity, the effect is diminished (Gaesser, Dodds, & Schacter, 2017; Sawczak, McAndrews, Gaesser, & Moscovitch, 2019).

These effects of episodic memory on self-reported prsosional intentions are related to phenomenological qualities of the recollected events, such as the amount of sensory detail that people can remember, or how coherent the recollected scene appears in one’s mind (e.g., Gaesser & Schacter, 2014, Exp. 3), both of which are considered measures of vividness (e.g., D’Argembeau & Van der Linden, 2012; see also Arnold, McDermott, & Szpunar, 2011; Hassabis, Kumaran, Vann, & Maguire, 2007). Other work has found that people tend to be more confident in the veracity of their memories when they are more vivid (Rubin, Schrauf, & Greenberg, 2003; Talarico & Rubin, 2007), which would seem to be important when making judgments about one’s own past behaviour, as in the present studies. People’s subjective ratings of the degree or vividness of their episodic memories have been reliably associated with neural activity in the hippocampus and MTL more broadly (e.g., Addis, Moscovitch, Crawley, & McAndrews, 2004; Addis, Roberts, & Schacter, 2011; Addis & Schacter, 2008; Gilboa, Winocur, Grady, Hevenor, & Moscovitch, 2004; Thakral, Madore, & Schacter, 2020), and people with compromised MTL functioning, whether due to natural aging or injury, tend to have less detailed episodic memories (e.g., Addis, Musicaro, Pan, & Schacter, 2010; Levine et al., 2002; Sheldon, McAndrews, & Moscovitch, 2011; St-Laurent et al., 2014). In the present studies, we expected that the vividness of participants’ episodic memories would be related to their TEQ item ratings, and that participants with compromised MTL functioning would report less vivid memories.

Although self-report trait questionnaires, such as the TEQ, tend not to ask respondents explicitly to retrieve memories while completing them, people may do so spontaneously, as cues in the item text on such questionnaires may elicit involuntary retrieval, a common occurrence in everyday life (Robin, Garzon, & Moscovitch, 2019; for review see Berntsen, 2007). Therefore, focusing on spontaneous retrieval as we did in our first three studies may afford a more ecologically valid study of the role of episodic memory in informing self-knowledge. In the last of these experiments presented here, we do explicitly manipulate retrieval, as when older adults and people with MTL excisions engage in a similar activity, the effect is diminished (Gaesser, Dodds, & Schacter, 2017; Sawczak, McAndrews, Gaesser, & Moscovitch, 2019).

In the four studies presented here, we do explicitly manipulate retrieval, as when older adults and people with MTL excisions engage in a similar activity, the effect is diminished (Gaesser, Dodds, & Schacter, 2017; Sawczak, McAndrews, Gaesser, & Moscovitch, 2019).

In Experiment 1, we introduced a novel approach called the Spontaneous Memory-Identity Survey (SMIS). First, a group of healthy, young adults filled out the TEQ and then, immediately afterward, were (unexpectedly) given the SMIS. In the SMIS, participants were asked to indicate whether, for each item on the TEQ, they had spontaneously recalled a specific memory related to the item when they were doing the initial rating. The resulting data were used to classify each item as either having cued a specific, autobiographical memory congruent with the behaviour measured by that item, or not. This allowed us to compare the ratings of these two categories of items, and our prediction was that participants would identify more strongly with (i.e., give higher ratings to) items that cued a relevant episodic memory. We also predicted that, among such items, ratings would correlate with the vividness (sensory detail and scene coherence) of participants’ memories.

In Experiment 2, we extended our investigation by comparing young and older adults in the same paradigm. In light of the research reviewed above, we predicted that older adults would report fewer and less vivid episodic memories compared to young adults, and for those memories they did retrieve, their ratings on TEQ items would be less influenced by episodic memory retrieval compared to those of young adults.

In Experiment 3, again using the same method, we compared individuals who had had MTL excisions (mTLE, as treatment for intracerebral epilepsy) to healthy, age-matched controls. We predicted that the mTLE cohort would report fewer and less vivid episodic memories, and that their TEQ item ratings would be less influenced by episodic memory retrieval compared to the control group.

Finally, in Experiment 4, we modified our method to compare the effects of spontaneous versus directed retrieval on TEQ item ratings. This study was conducted in healthy young adults only, and we predicted that directed retrieval would amplify ratings on their associated TEQ items, just as spontaneous retrieval does.

2. Statistical power and analytic approach

2.1. Statistical power

As we were unaware of any directly comparable studies at the time these experiments were conducted, the sample sizes for the present Experiments 1–3 were not informed by a priori power analyses. One study that we have since become aware of, however, and cited above (Charlesworth et al., 2016, Exp. 1), is comparable to the present Experiment 1 methodologically in that it tested for—and observed—a direct relationship between episodic autobiographical memory retrieval and reported self-knowledge. The relevant effect size for that study was $\eta_p^2 = 0.13$ (a medium-large effect; Cohen, 1969; Richardson, 2011), and a power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) shows that 24 participants is sufficient to detect an effect of that size with 95% power. With a slightly lower sample size of 23, a post-hoc power analysis finds that our Experiment 1 still had over 90% power to detect such an effect.

Similarly, our sample sizes in Experiment 2 were large enough to detect an effect of that size in each of the young and older adult groups separately (N’s = 25 and 26, respectively), with over 95% power. While our sample sizes in Experiment 3 were admittedly modest (N’s = 15 mTLE participants and 15 control participants), they are considered reasonable for studies of autobiographical memory among individuals with unilateral MTL damage. A post-hoc power analysis showed that Experiment 3 had almost exactly 80% power to detect the effect of interest in each of the mTLE and control groups separately.

The sample sizes for Experiment 4 (N = 100 in each group of participants; see our pre-registration submission at https://aspredicted.org/us7np.pdf) were partially informed by a power analysis of the between-participant effect sizes, averaged across Experiments 2 (young vs. older adults) and 3 (mTLE patients vs. healthy controls), for the mean difference in ratings of TEQ items that cued an episodic memory (Cohen’s $d = 0.653$, a medium-sized effect; Cohen, 1988). While the power analysis called for 62 participants in each group, we took a conservative approach and recruited 100 in each group (before analyzing any of the data), given that the study was being conducted remotely (online) for the first time, and we were unsure how sensitive our experimental design would be in such a context.

2.2. Analytic approach

We adopted a multi-level modeling approach to contend with missing data for participants. In this approach, individual TEQ items (nested within participants) are the unit of analysis, as opposed to...
participant means. Model specification details are provided in the Results section for each experiment. In all analyses, our threshold for statistical significance was $\alpha = 0.05$, though in some cases we draw attention to trending results ($\alpha = 0.1$). In an earlier version of this paper included in a dissertation (Sawczak, 2020), Experiments 1–3 were analyzed based on participant-level means, and those analyses yielded the same general pattern of results as did multi-level modeling.

3. Experiment 1: method

3.1. Participants

We recruited 25 young adults ($M_{\text{age}} = 20.7$, $SD = 3.61$, range = 18–34; 18 women) from the University of Toronto community, but excluded two participants who did not report a single episodic memory on the SMIS (final sample $N = 23$; $M_{\text{age}} = 20.10$, $SD = 2.26$, range 18–26; 16 women). All participants were fluent in English and had no history of psychological or neurological disorder. They gave written consent and were compensated either with $10 or with partial course credit if they were enrolled in an introductory psychology course. Our experimental protocol was approved by the University of Toronto Research Ethics Board.

3.2. Materials and procedure

3.2.1. Toronto Empathy Questionnaire

The Toronto Empathy Questionnaire (TEQ) is a 16-item self-report questionnaire that was created using a data-driven approach to identify a common empathy factor shared by various pre-existing measures of empathy (Spreng et al., 2009). As such, it contains items from the Interpersonal Reactivity Index (Davis, 1983), Hogan's Empathy Scale (Hogan, 1969), the Balanced Emotional Empathy Scale (Mehrabian, 2000), and others. The TEQ has been shown to have high internal consistency (Cronbach’s $\alpha = 0.85$) and test-retest reliability ($r = 0.81$, $p < .001$), and it is correlated ($r = 0.35$, $p < .01$) with the widely used “Reading the Mind in the Eyes” Test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), a behavioural measure of social inference. The TEQ includes both positive-scored (e.g., “I enjoy making other people feel better”; “I get a strong urge to help when I see someone who is upset”) and negatively-scored (e.g., “I become irritated when someone cries”; “When I see someone being treated unfairly, I do not feel very much pity for them”) items. Participants rated each item in terms of behavioural frequency (“Please read each statement carefully and rate how frequently you feel or act in the manner described”) on a 5-point Likert-type scale ranging from 0 (Never) to 4 (Always).

3.2.2. Spontaneous Memory-Identity Survey

We devised a 16-page ‘Spontaneous Memory-Identity Survey’ (SMIS) for the TEQ, with each page containing the text of a TEQ item followed by these instructions:

Did you think about a memory from your own life when you were making a rating for the above statement on the survey you were filling out earlier? If you did not, that is okay. You do not have to think of one now; you can simply skip to the next page of the survey.

Participating first answered the above question by checking one of two boxes labelled ‘yes’ and ‘no’. If participants responded ‘yes’, then they would answer the next question, ‘In the event you remembered, how were you acting?’ by checking one of two boxes labelled ‘I acted in a way that fits with the statement’ and ‘I acted in a way that does NOT fit with the statement’. Finally, participants rated two aspects of the vividness of the memory in question (sensory detail: 1 = vague, 7 = detailed; and scene coherence: 1 = fragmented, 7 = coherent), wrote a brief description of the memory, and moved on to the next page (next item). Importantly, they were not able to view their original TEQ responses while they were completing the SMIS.

After participants completed the SMIS, the experimenter sat down with them to review their written responses and asked whether each reported memory referred to a specific, unique event (i.e., episodic memory) or to a more general, repeated experience (i.e., semantic memory). In some cases, this procedure revealed that participants initially had some misunderstanding about what sort of memory qualified as valid to report on the SMIS, and their oral responses during this phase were used to confirm which TEQ items cued episodic memories. Such revisions occurred more often when participants’ written responses contained few words.

4. Experiment 1: results

All positive-scored TEQ items (e.g., “It upsets me to see someone being treated disrespectfully”) that reportedly cued an episodic memory congruent with the behaviour described by the item (61 items) were classified as “empathic memory items”. All reverse-scored items (e.g., “I remain unaffected when someone close to me is happy”) that reportedly cued an episodic memory incongruent with the item (and therefore demonstrative of empathic behaviour) were also included in this category (41 items). All other items, including but not restricted to those that did not cue an episodic memory at all, were classified as “all other items” (266 items; 368 in total). Next, we took a multi-level modeling approach where individual TEQ items, nested within participants, were the units of analysis. We ran a model of this sort with TEQ item rating as the dependent variable, a dummy-coded predictor variable for item type (1 = empathic memory items; 0 = all other items), and a random intercept based on participant ID. The model revealed a significant effect of item type wherein empathic memory items were rated more highly compared to all other items (see Table 1 for model output; see Fig. 1 for plot).

Next, to test the effect more stringently, we created a third category of items: “anti-empathic memory items”. This category contained positive-scored items for which an incongruent memory had been cued (i.e., demonstrative of anti-empathic behaviour) as well as reverse-scored items for which a congruent memory had been cued (also demonstrative of anti-empathic behaviour). These 24 items had previously been collapsed into the “all other items” category; our focus had been on memories of empathic behaviour and we did not originally anticipate these types of memories. We reasoned, however, that compared to items that did not cue a memory at all (“non-memory items”), anti-empathic memory items should be rated lower on average. To test this prediction, we ran a new model similar to the one specified above but comparing only anti-empathic memory and non-memory items. As predicted, anti-empathic memory items were rated significantly lower than non-memory items ($B = -0.392$, $SE = 0.170$, 95% CI = [−0.658 to −0.126], $t_{(229)} = 3.014$, $p < .01$, model marginal $R^2 = 0.039$). We then ran a third model comparing only empathic memory items and non-memory items, now that they had been separated from anti-empathic memory items, and again found that empathic memory items were rated significantly higher ($B = 0.391$, $SE = 0.089$, 95% CI = [0.215–0.567], std. Beta = 0.485, $t_{(320)} = 4.578$, $p < .001$, model marginal $R^2 = 0.088$).

The average sensory detail rating for all episodic memories across all participants was 4.71 out of 7 ($SD = 1.57$) and the average scene coherence rating was 4.60 out of 7 ($SD = 1.79$). We ran one last multi-level model (using only empathic memory items) to test our prediction that vividness (sensory detail and scene coherence ratings) would predict higher ratings on empathic memory items, but did not detect a significant effect of sensory detail ($B = 0.019$, $SE = 0.056$, 95% CI = [−0.211–0.301], std. Beta = 0.045, $t_{(277)} = 0.349$, $p = .728$, model marginal $R^2 = 0.001$) or of scene coherence ($B = -0.008$, $SE = 0.052$, 95% CI = [−0.112–0.095], std. Beta = −0.022, $t_{(277)} = -0.159$, $p = .874$).
knowledge can be informed by episodic memory, at least in some circumstances, consistent with other work (e.g., Charlesworth et al., 2016; Wolf & Zimprich, 2016), and raise the question of whether people with impairments in episodic memory and MTL function would show a diminished relationship between episodic memory and self-knowledge.

6. Experiment 2

Older age is associated with a decline in episodic memory (for reviews see Hedden & Gabrieli, 2004; Park & Gutchess, 2005), specifically regarding level of episodic detail (Levine et al., 2002; Piolino et al., 2010) and “autonoetic consciousness”, which is the subjective sense of mentally traveling back in time and re-experiencing a memory while recollecting it (Piolino et al., 2006; Tulving, 1985). Thus, we might expect older adults to report fewer episodic memories being cued, and this might result in an attenuated relationship between episodic recall and TEQ item ratings compared to young adults. In addition, compared to older adults, young adults tend to have less life experience and are in the formative stages of personality development (Cramer, 2017; Erikson, 1950; Kroger et al., 2010; Marcia, 1967). Therefore young adults may search for and rely on episodic “evidence” to infer how frequently they manifest the empathic behaviours described by the TEQ, whereas older adults may rely on crystallized, semantic knowledge of their own personalities that has already been abstracted from individual episodes (e.g., Moscovitch et al., 2005; Umanath & Marsh, 2014). Thus, we might expect the SMIS to show that TEQ items that cue episodic memories are more strongly endorsed by young adults but that this effect is diminished in older adults. Since we also wished to replicate the SMIS effect we found in Experiment 1, for Experiment 2 we recruited a second sample of young adults in addition to a sample of older adults.

7. Experiment 2: method

The method of Experiment 2 was identical to that of Experiment 1 aside from the participants involved.

7.1. Participants

We recruited 25 young adults (M<sub>age</sub> = 19.08, SD = 3.51; range = 17–34, 18 women) and 26 older adults (M<sub>age</sub> = 73.42, SD = 4.70, range = 65–81; 18 women) from the University of Toronto community. As in Experiment 1, all participants were fluent in English, had no history of psychological or neurological disorder and gave written consent. Young adults were compensated either with $10 or with course credit if they were enrolled in an introductory psychology course; older adults were compensated with $16. Our experiment protocol was approved by the University of Toronto Research Ethics Board.

8. Experiment 2: results

First, we compared young and older adults on their total TEQ score and found no significant difference between the groups (young adults M = 49.6, SD = 6.39; older adults M = 46.77, SD = 7.21), t<sub>49</sub> = 1.48, p = .145, d = 0.416. Next, TEQ items were categorized in the same way as in Experiment 1 (young adults: 111 empathic memory items, 289 other items; older adults: 90 empathic memory items, 326 other items) and a multi-level model was conducted. The model was specified in the same way as in Experiment 1 with the addition of a predictor for age group (dummy coded: 1 = young adults; 0 = older adults) and an interaction
term between age group and item type. Critically, the model revealed a significant interaction (see Table 2 for model output; see Fig. 2 for plot).

Given the interaction effect, we ran a post-hoc model for each age group separately, specified as above but with item type as the only predictor. The model for young adults showed a significant effect of item type, wherein empathic memory items were rated more highly (B = 0.420, SE = 0.089, 95% CI = [0.246–0.595], std. Beta = 0.485, t(374) = 4.727, p < .001, model marginal R² = 0.047), but there was no significant effect of item type in the model for older adults (B = 0.112, SE = 0.097, 95% CI = [–0.079–0.303], std. Beta = 0.126, t(389) = 1.155, p = .249, model marginal R² = 0.003).

Since there were so few anti-empathic memory items in each age group (young adults: 15 items; older adults: 9 items; see Supplementary Material for frequencies of item types across all groups and studies), we opted not to analyze them separately. We did, however, rerun the models above after excluding anti-empathic memory items from the “all other items” category, and found an identical pattern of results: The full model showed no simple effects (item type: B = 0.099, SE = 0.094, 95% CI = [–0.085–0.283], std. Beta = 0.116, t(739) = 1.061, p = .289; age group: B = 0.115, SE = 0.120, 95% CI = [–0.126–0.355], std. Beta = 0.134, t(49) = 0.960, p = .342, model marginal R² = 0.032) but a significant interaction between item type and age group (B = 0.261, SE = 0.129, 95% CI = [0.008–0.515], std. Beta = 0.306, t(739) = 2.026, p = .043), and the post-hoc models showed a significant effect of item type for young adults (B = 0.362, SE = 0.086, 95% CI = [0.193–0.532], std. Beta = 0.439, t(739) = 4.199, p < .001, model marginal R² = 0.040) but not for older adults (B = 0.099, SE = 0.096, 95% CI = [–0.090–0.288], std. Beta = 0.113, t(389) = 1.030, p = .303, model marginal R² = 0.002).

Next, we checked whether the age groups differed in the mean number of empathic memory items they reported, expecting older adults to report significantly fewer of them than young adults, as per well-established differences in episodic memory retrieval (e.g., Addis et al., 2010; Levine et al., 2002). A t-test showed that older adults reported numerically fewer memories compared to young adults, but the difference was only trending and not significant (young adults M = 4.44, SD = 1.83; older adults M = 3.50, SD = 1.73), t(49) = –1.89, p = .065, d = 0.529. Further, the age groups did not differ significantly in terms of sensory detail ratings (young adults M = 4.99, SD = 1.46; older adults M = 5.52, SD = 0.89), t(49) = 1.57, p = .123, d = 0.437, or scene coherence ratings (young adults M = 4.96, SD = 1.52; older adults M = 5.17, SD = 1.28), t(49) = 0.54, p = .590, d = 0.152. To test our prediction that these measures of vividness would be related to item ratings, we ran two more multi-level models, for young and older adults separately, with item rating as the dependent variable, sensory detail and scene coherence as predictors, and a random intercept based on participant ID. Only empathic memory items were included in these models. The model for young adults showed that sensory detail predicted significantly higher item ratings (B = 0.149, SE = 0.069, 95% CI = [0.012–0.286], std. Beta = 0.330, t(70) = 2.163, p = .034, model marginal R² = 0.054), but scene coherence did not (B = –0.072, SE = 0.064, 95% CI = [–0.201–0.056], std. Beta = –0.171, t(70) = –1.124, p = .265). The model for older adults did not show a significant effect of sensory detail (B = 0.087, SE = 0.126, 95% CI = [–0.166–0.340], std. Beta = 0.118, t(58) = 0.685, p = .496, model marginal R² = 0.012) nor of scene coherence (B = –0.005, SE = 0.096, 95% CI = [–0.198–0.188], std. Beta = –0.009, t(58) = –0.051, p = .959). However, given that the confidence intervals for the effects of sensory detail in young and older adults overlap, we suggest the relationship between sensory detail and item ratings is numerically smaller among older adults, but not significantly so.

9. Experiment 2: discussion

Young adults gave significantly higher ratings to TEQ items that reportedly cued episodic memories, while older adults did not. These item ratings were predicted by sensory detail ratings in young adults to a slightly greater degree than they were in older adults. Regarding the young adults, it is noteworthy that while the first finding replicates Experiment 1, the second finding was not observed in the earlier study, despite having a nearly identical sample of young adults. Given that the methodology was identical between the two experiments, potential explanations for the discrepancy are elusive, but the upcoming two experiments may be instructive in this regard. Our original thinking was that some aspect of memory vividness—whether the richness of specific details, the coherence of the spatial context, or both—would be positively related to self-reported behaviour frequency, given that people consider more vivid memories to be more veridical (Rubin et al., 2003;
specific episodic memories when rating their own empathy because they have intractable epilepsy. Ever since the famous case of patient H.M. it has been known that such excisions result in severe impairments of episodic memory (Penfield & Milner, 1958; Squire, 2009). Typically, however, excisions are unilateral, being confined to the hemisphere in which seizures originate. The resulting memory impairments are far less severe, yet verifiable (e.g., Baxendale, Thompson, & Sander, 2013; Lee, Yip, & Jones-Gotman, 2002; McAndrews, 2012). In particular, when reporting autobiographical memories, individuals with mTLEs generate significantly fewer episodic (context-specific) details compared to healthy controls (St-Laurent et al., 2009; Viskontas, McAndrews, & Moscovitch, 2000). Particularly relevant to the present work, one study found that people with MTL epilepsy, even without having had excisions, reported fewer autobiographical memories compared to healthy controls, and that this was associated with abnormal scores on a measure of personal identity (Allebone et al., 2015). Therefore, we expected that, compared to healthy controls, individuals with mTLEs would report fewer and less vivid episodic autobiographical memories in response to items on the TEQ, and that items associated with such memories would not be rated significantly higher than the other items.

11. Experiment 3: method

The method of Experiment 3 was identical to that of Experiment 2 aside from the participants involved.

11.1. Participants

Through the epilepsy clinic at Toronto Western Hospital, we recruited 16 individuals (11 women; M_age = 37.6, SD = 11.7, range = 22–60; M_years of education = 15.3, SD = 2.3) who had undergone unilateral anterior temporal lobectomy for treatment of intractable epilepsy at least six months prior and had been seizure-free since. The excisions included the entirety of the amygdala, hippocampus, and anterior parahippocampal gyrus as well as 4–5 cm from the inferior temporal gyrus, 2–3 cm from the middle temporal gyrus, and 1–2 cm from the superior temporal gyrus. While eight of the patients’ excisions were in the left hemisphere and eight were in the right, previous work from our lab has found that laterality is not a significant factor when it comes to autobiographical memory deficits (Seldon et al., 2011; St-Laurent et al., 2009; Viskontas et al., 2000). These participants were compensated with coffee shop gift cards worth $25, according to a pre-established policy in our lab at Toronto Western Hospital. Through the University of Toronto community, we also recruited 16 healthy control participants matched to the patients, on an individual basis, for age, sex, and years of education (11 women; M_age = 37.1, SD = 11.7, range = 23–61; M_years of education = 16.3, SD = 2.7), and with no history of psychiatric or neurological conditions. These participants were compensated with $16, in accordance with the standards of our lab at the University of Toronto. The experiment protocol was approved by the University Health Network Research Ethics Board.

12. Experiment 3: results

One person in the mTLE group and one person in the control group did not report a single episodic memory on the SMIS, so we excluded these two individuals from analysis. First, we compared the two groups on total TEQ score and found no significant difference (mTLE M = 48.3, SD = 6.2; control M = 49.5, SD = 5.0), t(28) = 0.55, p = .586, d = 0.201. There was also no significant group difference in sensory detail ratings (mTLE M = 5.5, SD = 1.2; control M = 5.5, SD = 1.2), t(28) = 0.09, p = .929, d = 0.033, or scene coherence ratings (mTLE M = 5.0, SD = 1.1; control M = 5.4, SD = 1.2), t(28) = 1.01, d = 0.320, d = 0.370, similar to Experiment 2. As has been suggested in the case of older adults (Addis et al., 2011), it could be that when mTLE patients make such sorts of subjective ratings, they reflect general knowledge more than episodic content. The mTLE group reported slightly fewer episodic memories (M = 3.4, SD = 2.1) than the control group (M = 3.9, SD = 1.8), but this...
difference was not significant, \( t_{(28)} = 0.653, p = .519 \).

After categorizing the different types of items, we ran a multi-level model specified in the same way as in Experiment 2, with TEQ item rating as the dependent variable, a predictor for item type (1 = empathic memory items; 0 = all other items), a predictor for participant group (1 = controls; 0 = mTLEs), an interaction term, and a random intercept based on participant ID. There were no significant effects of item type (B = 0.105, SE = 0.135, 95% CI = [–0.159–0.370], std. Beta = 0.121, \( t_{(448)} = 0.782, p = .434 \), model marginal \( R^2 = 0.012 \)) or participant group (B = 0.027, SE = 0.133, 95% CI = [–0.246–0.300], std. Beta = 0.031, \( t_{(28)} = 0.201, p = .842 \), nor was there a significant interaction (B = 0.170, SE = 0.185, 95% CI = [–0.195–0.534], std. Beta = 0.194, \( t_{(448)} = 0.915, p = .361 \); see Fig. 3 for plot).

Given the modest sample size, however, and the fact that we had a clear a priori hypothesis that the SMIS effect would be stronger in the control group versus the mTLE group, we decided to examine the groups separately using the same model above (without, of course, terms for a simple effect of participant group and its interaction with item type). This analysis revealed that although the effect of item type remained non-significant for the mTLE group, it was indeed significant for the control group (see Table 3A and 3B for model output).

As in Experiment 2, we chose not to analyze anti-empathic memory items separately because there were so few of them (controls: 6 items; mTLEs: 6 items; see Supplementary Material for frequencies of item types across all groups and studies), but we did rerun the models above after excluding anti-empathic memory items from the “all other items” category. The pattern of results was similar: In the full model there was no simple effect of item type (B = 0.082, SE = 0.132, 95% CI = [–0.178–0.341], std. Beta = 0.096, \( t_{(436)} = 0.618, p = .537 \), model marginal \( R^2 = 0.010 \)), nor of participant group (B = 0.038, SE = 0.132, 95% CI = [–0.232–0.308], std. Beta = 0.044, \( t_{(28)} = 0.286, p = .777 \), and no interaction between item type and participant group (B = 0.155, SE = 0.182, 95% CI = [–0.202–0.512], std. Beta = 0.182, \( t_{(436)} = 0.853, p = .394 \)). Post-hoc models, however, again showed a significant effect of item type for controls (B = 0.241, SE = 0.113, 95% CI = [0.017–0.464], std. Beta = 0.312, \( t_{(218)} = 2.122, p = .035 \), model marginal \( R^2 = 0.018 \)) but not for mTLEs (B = 0.081, SE = 0.143, 95% CI = [–0.201–0.362], std. Beta = 0.088, \( t_{(218)} = 0.564, p = .573 \), model marginal \( R^2 = 0.001 \)).

Next, to test our prediction that the vividness (sensory detail and scene coherence) of memories would be related to item ratings, we ran two more multi-level models, for controls and mTLEs separately (younger samples), specified in the same way as the analogous models in Experiment 2. The model for controls showed that neither sensory detail predicted significantly higher item ratings (B = 0.048, SE = 0.125, 95% CI = [–0.300–0.204], std. Beta = –0.079, \( t_{(40)} = –0.387, p = .701 \), model marginal \( R^2 = 0.016 \)), nor scene coherence predicted significantly higher item ratings (B = 0.093, SE = 0.107, 95% CI = [–0.123–0.308], std. Beta = 0.177, \( t_{(40)} = 0.870, p = .390 \)). The model for mTLEs did not show a significant effect of sensory detail (B = –0.111, SE = 0.137, 95% CI = [–0.390–0.167], std. Beta = –0.186, \( t_{(33)} = –0.814, p = .422 \), model marginal \( R^2 = 0.119 \)) but it did show a significant effect of scene coherence (B = 0.268, SE = 0.127, 95% CI = [0.010–0.527], std. Beta = 0.482, \( t_{(28)} = 2.112, p = .042 \)). Similar to Experiment 2, however, the confidence intervals for the effects of scene coherence in mTLEs and controls overlap, suggesting these effects are numerically different but not significantly so.

Last, given our age-related findings in Experiment 2, we deemed it worth considering how a potential aging effect in the present study might interact with or mask the effect of mTLEs. While the mean age of the mTLE group (M = 38.0, SD = 12.0) was not significantly different from that of the control group (M = 38.0, SD = 12.1), \( t_{(28)} = 0.005, p = .996, d = 0.002 \), over half of the participants in each group were well over the age of 35, which is the standard cut-off for “young adult” cohorts in our lab (see also Conway, Wang, Hanuy, & Haque, 2005; Cramer, 2017; Frebble et al., 2013).

In our initial model reported above, predicting TEQ item rating with item type and participant group, we did not find a significant interaction between item type and participant group, nor a simple effect of either variable. It was only once we modeled participant groups separately that we found an effect of item type in the control group but not in the mTLE group. This is likely due to the small sample size in the present study which, while common in studies of autobiographical memory in people with MTL damage, is likely underpowered to detect significant interactions. Hence, an even more sophisticated model, allowing for a three-way interaction between age, participant group, and item type, would be even less likely to detect significant effects (especially since the age range is narrower compared to Experiment 2). For completeness, however, we ran such a model, using a categorical variable for age group (age 35 and under vs. over age 35). As expected, there were still no interactions involving item type.

For a final exploratory analysis, we modeled only participants aged 35 and under (reducing our sample size by half), reasoning that any memory-related difference between the control and mTLE groups should be more strongly manifest among young adults. The omnibus model was now able to detect a trending interaction between item type and participant group, and post-hoc models for each participant group again showed a significant effect of item type in controls but not in mTLEs (see Supplementary Material). With this reduced sample of young adults only, we also found that sensory detail significantly predicted higher item ratings in the control group, while neither sensory detail nor scene coherence had a significant effect in the mTLE group (see Supplementary Material).

13. Experiment 3: discussion

In Experiment 3 we initially did not find our predicted results regarding episodic memory retrieval and its different effects in neurologically healthy control participants versus those with mTLEs. We suspected there were multiple reasons for this, including the small sample size and the fact that some of the healthy controls were older than the young adults in Experiments 1 and 2. Therefore, we examined the control group and mTLE group separately, and found our primary effect of interest: Healthy controls gave significantly higher ratings to TEQ items that reportedly cued episodic memories, while their counterparts who had mTLEs did not. These item ratings were predicted by scene coherence ratings a little more strongly in the mTLE group compared to the control group, though in an exploratory analysis of only younger participants, sensory detail predicted item ratings in healthy controls, while neither sensory detail nor scene coherence had an effect in the mTLE group (see Supplementary Material).

Although the findings from Experiment 3 should be taken with caution, they suggest that MTL excisions interfere with the SMIS effect. Interestingly, another group that compared individuals with medial temporal lobe epilepsy to people whose seizures originated elsewhere in...
with a study by Schwarz et al. (1991), who found that when instructions or a given item has some influence over TEQ ratings. This notion dovetails with graphical memories and judgments about one's trait empathy.

D'Argembeau and Van der Linden (2010) discussed multiple components to vividness and measured them separately in addition to overall vividness (see also Talarico & Rubin, 2007). Others have found effects of episodic memory on decision-making to be modulated by vividness ratings (e.g., Peters & Büchel, 2010). Given the mixed results regarding sensory detail and scene coherence ratings in Experiments 1–3, asking participants to rate the vividness of memories cued by TEQ items may be a more reliable way of getting at the relationship between the episodic content of autobiographical memories and judgments about one’s trait empathy.

In addition, it is conceivable that the failure to retrieve a memory for a given item has some influence over TEQ ratings. This notion dovetails with a study by Schwarz et al. (1991), who found that when instructions required participants to come up with more memories than they easily could, participants rated themselves lower on the trait adjective in question. In other words, the failure of retrieval may be just as important as success, and the perceived difference between these two types of subjective experiences may be important while one is making relevant judgments (Schwarz & Strack, 2016). While not the primary aim of Experiment 4, our manipulation of directed versus spontaneous retrieval may shed some light on this question as well.

### 14. Experiment 4

Our findings in Experiments 1–3 generally show that episodic memory retrieval is associated with higher ratings on TEQ items, at least among younger and neurologically healthy adults. Importantly, however, in those experiments participants were asked after the fact which items spontaneously cued episodic memories. It is conceivable that participants may have either remembered that they gave a relatively high rating to certain items beforehand or identified strongly with the behaviours described in those items (these two possibilities are not mutually exclusive) and, therefore, were more inclined to search for a relevant memory due to demand characteristics. Experiment 4 was designed to control for this possibility.

### 15. Experiment 4: method

#### 15.1. Participants

Participants were recruited through Amazon’s Mechanical Turk online community. Originally, we performed an a priori power analysis to determine our sample size (see Statistical Power and Analytic Approach section), which called for that 62 participants per condition. We aimed, however, to take a conservative approach, given potential unforeseen differences in executing our methodology in a remote (online) setting.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.999</td>
<td>0.104</td>
<td>-0.024</td>
<td>2.793–3.204</td>
<td>28.704</td>
<td>&lt; 0.001</td>
<td>224</td>
</tr>
<tr>
<td>Item type</td>
<td>0.105</td>
<td>0.144</td>
<td>0.112</td>
<td>-0.180–0.389</td>
<td>0.725</td>
<td>0.469</td>
<td>224</td>
</tr>
</tbody>
</table>

Random Effects

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>σ²</td>
<td>0.781</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>τ0 pid</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC</td>
<td>0.114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N pid</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal R²</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditional R²</td>
<td>0.116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Item type was dummy-coded (1 = empathic memory items; 0 = all other items).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.024</td>
<td>0.083</td>
<td>-0.083</td>
<td>2.862–3.187</td>
<td>36.623</td>
<td>&lt; 0.001</td>
<td>224</td>
</tr>
<tr>
<td>Item type</td>
<td>0.278</td>
<td>0.118</td>
<td>0.344</td>
<td>0.046–0.510</td>
<td>2.599</td>
<td>0.019</td>
<td>224</td>
</tr>
</tbody>
</table>

Random Effects

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>σ²</td>
<td>0.588</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>τ0 pid</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC</td>
<td>0.083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N pid</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal R²</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditional R²</td>
<td>0.103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Item type was dummy-coded (1 = empathic memory items; 0 = all other items).
and aimed for 100 participants per condition. After excluding 74 participants for failed attention checks or duplicate IP addresses, we were left with usable data from a total of 200 participants (control condition: \( N = 100, \text{M}_{\text{age}} = 35.7, \text{SD} = 10.9, \text{range} = 20–71, 34 \text{ women} \); retrieval condition: \( N = 100, \text{M}_{\text{age}} = 36.6, \text{SD} = 10.8, \text{range} = 21–65, 51 \text{ women} \)). Participants were compensated for their time at the rate of $5 per hour.

### 15.2. Materials and procedure

#### 15.2.1. Toronto Empathy Questionnaire

As in the preceding studies, we measured trait empathy using the Toronto Empathy Questionnaire. In this study, however, participants responded to each TEQ item on a scale of behavioural frequency from 1 (Never) to 5 (Always) as opposed to 0 to 4.

#### 15.2.2. Spontaneous Memory-Identity Survey

The Spontaneous Memory-Identity Survey (SMIS) in Experiment 4 was administered online, instead of on paper, to enable data collection through Amazon’s Mechanical Turk. Participants also received more detailed instructions about what constitutes an episodic memory (see Supplementary Material). The only other difference, compared to Experiments 1–3, was that when participants indicated that an item cued a memory, they were also asked to indicate how recent the memory was (1 = past week; 2 = past month; 3 = past year; 4 = past 10 years; 5 = longer), and they rated the overall vividness of the memory (1 = least vivid; 4 = most vivid) instead of sensory detail and scene coherence. Memory recency was measured for exploratory purposes and is not analyzed here.

Participants were randomly assigned to one of two conditions in a between-subjects design: a control condition and a retrieve condition. In the control condition, participants answered the entire Toronto Empathy Questionnaire (TEQ), then completed an adapted version of the SMIS (see Measures) asking about spontaneous memory recall for each TEQ item. In the retrieve condition, each TEQ item was preceded by explicit instructions to recall a specific memory from one’s own life in which one was behaving in a way that was congruent with the item text. Further, in the retrieve condition, participants answered SMIS questions for each TEQ item immediately after responding to the item.

Since Experiment 4 was being conducted remotely, participants in both conditions also read detailed instructions (see Supplementary Material) explaining what constitutes an episodic memory and that only episodic memories congruent with the item text should be reported on the SMIS. These instructions were followed by practice questions in which participants were shown an example item statement not contained in the TEQ (e.g., “I like to go fishing”), along with a brief description of a memory, and had to indicate whether the memory would be appropriate to report on the SMIS based on its episodic quality and its congruency with the item text. The instructions and practice questions were shown to participants immediately before they began the SMIS, whether at the beginning of the experiment (for participants in the retrieve condition) or after completing the TEQ (for participants in the control condition).

### 16. Experiment 4: results

An important difference in our analytic approach to Experiment 4, as compared to the previous experiments, is that we analyzed positive- and reverse-scored items separately. In Experiments 1–3, participants reported very few episodic memories cued by reverse-scored items (see Supplementary Material for frequencies of item types across all groups and studies). Nonetheless, memories that were “incongruent” with the behaviour described by reverse-scored item statements, and therefore demonstrative of empathic behaviour, were collapsed into the “empathic memory” category. Memories that were incongruent with positive-scored items were referred to as “anti-empathic memory items” and were collapsed into the “all other items” category for our primary analyses, though we also ran auxiliary analyses to see whether excluding them altogether altered our results (it did not). In Experiment 4, however, we obtained many more of these incongruent items, likely because of the larger sample size and the fact that in the Retrieve condition participants were instructed to retrieve a congruent memory for every item. Therefore, we decided it would be worthwhile to treat reverse-scored items separately in all analyses. In addition, to capitalize on the fact that the vast majority of memories in Experiment 4 were congruent with the item text, regardless of the item being positive- or reverse-scored, we decided to exclude from analysis altogether the small number of incongruent memories.

Before running any models, we first removed items for which participants indicated that they had recalled an incongruent memory. If enough participants had reported an incongruent memory, we would have analyzed these items separately, but only 41 memories out of 1257 total recalled memories between both conditions involved an incongruent memory, which did not allow for analysis of such items. We then divided the data set so that we could analyze positive- and reverse-scored items separately.

Next, we tested a model for each subset of data with item rating as the dependent variable; memory retrieval (1 = success; 0 = failure) and experimental condition (1 = retrieve; 0 = control) as predictors; an interaction term; and a random intercept based on participant ID.

For positive-scored items, successful memory retrieval predicted significantly higher item ratings; being in the retrieve condition predicted significantly lower item ratings; and there was a significant interaction between the two variables (see Table 4A for model output; see Fig. 4 for plot). Post-hoc testing revealed that the interaction effect occurred because items that did not cue a memory were rated significantly lower in the retrieve condition (\( \beta = -0.792, SE = 0.142, 95\% CI = [-1.073 \text{ to } -0.512], \text{std. Beta} = -0.716, t_{(160)} = -5.588, p < .001 \)), while items that did cue a memory did not differ by condition (\( \beta = -0.122, SE = 0.105, 95\% CI = [-0.330 \text{ to } -0.085], \text{std. Beta} = -0.151, t_{(150)} = -1.163, p = .247 \)).

In the model for reverse-scored items, successful memory retrieval predicted significantly lower item ratings; being in the retrieve condition predicted significantly lower item ratings; and there was a significant interaction between the two variables (see Table 4B for model output; see Fig. 4 for plot). Post-hoc tests showed that the interaction effect arose because items that did not cue a memory were rated significantly higher in the retrieve condition (\( \beta = 0.752, SE = 0.106, 95\% CI = [0.543 \text{ to } 0.962], \text{std. Beta} = 0.757, t_{(176)} = 7.087, p < .001 \)) but items that did cue a memory did not differ by condition (\( \beta = 0.061, SE = 0.150, 95\% CI = [-0.236 \text{ to } -0.359], \text{std. Beta} = 0.069, t_{(122)} = 0.408, p = .684 \)), mirroring the results for positive-scored items.

We then divided the data again, this time so that we could examine the control condition in isolation, in the interest of considering whether the data replicated findings from Experiments 1–3 regarding the effect of spontaneous memory recall on TEQ item ratings. Again, we analyzed reverse-scored and positive-scored items in separate models. We included memory retrieval as a predictor (1 = success; 0 = failure) in each model along with a random intercept for participant. These models revealed that successful memory retrieval predicted significantly higher ratings for positive-scored items, replicating Experiments 1–3, as well as significantly lower ratings for reverse-scored items (see Table 5A and 5B for model output).

Next, we examined whether explicitly directed retrieval, when successful at the time of item rating, had different effects compared to the absence of instructions to retrieve. In other words, we compared ratings for all items in the control condition, including those for which participants retrospectively said they had spontaneously retrieved a memory, to ratings for those items in the retrieve condition for which participants said they were able to retrieve a memory at the time when asked. We used the same modeling approach as above, again analyzing positive- and reverse-scored items separately. The models showed that positive-
scored items were rated significantly higher in the retrieve condition compared to the control condition, and that reverse-scored items were rated significantly lower in the retrieve condition compared to the control condition (see Table 6A and 6B for model output). These results indicate that explicitly directing episodic memory retrieval prior to responding to a TEQ item significantly influences the rating given on that item.

Finally, using the same modeling approach, we considered the role of vividness in TEQ item ratings. The following models involve only those items for which participants successfully retrieved a memory, and since we had found earlier that item ratings associated with successfully retrieved memories did not differ by condition, we excluded that variable from these models. Among positive-scored items, memory vividness predicted significantly higher ratings, and among reverse-scored items, it predicted significantly lower ratings (see Table 7A and 7B for model output; see Fig. 5 for plots).

It is worth noting that four participants in this experiment were at least 65 years old. We did not require participants to be within a certain age range for this experiment, as our aim was to test a broad sample of the general population. In light of our age-related findings in Experiment 2, however, we reran all of the above analyses excluding our four “senior” participants and obtained the same general pattern of results. Last, for exploratory purposes, we also reran the models above with participant age included as a covariate, and again obtained the same general pattern of results. There were no effects of age, likely because the vast majority of these participants were younger than the older adult cohort in Experiment 2.

17. Experiment 4: discussion

In Experiment 4 we explicitly manipulated whether episodic memory retrieval was required at the time when participants rated themselves on TEQ items. This design choice was made in order to address the issue that, in Experiments 1–3, participants were asked retrospectively whether they had spontaneous memories cued by TEQ items. We found...
that positive-scored items in the retrieve condition, for which participants recalled examples of themselves exhibiting empathic behavior prior to making a rating, were rated significantly higher on average compared to positive-scored items in the control condition. Likewise, reverse-scored items in the retrieve condition, for which participants recalled examples of themselves exhibiting non-empathic behavior prior to making a rating, were rated significantly lower on average compared to reverse-scored items in the control condition. Interestingly, however, when items in the retrieve condition associated with successful memory retrieval were compared to only those items in the control condition for which participants retrospectively said they had spontaneously recalled a memory, there were no significant differences. The fact that the ratings on items associated with episodic memory retrieval were similar for both directed and spontaneous retrieval lends credence to participants’ retrospective reporting of spontaneously retrieved memories not only in the present experiment but in Experiments 1–3. That is, there is reason to believe that when participants say they spontaneously retrieved a memory when first filling out the TEQ, they actually did.

It is important to note that it was not uncommon for participants in the retrieve condition to say they were unable to recall a memory when they were instructed to do so. There was a total of 588 retrieval failures out of 1600 items across 200 participants. Such items were rated significantly lower (or higher) than items that did not cue a memory in the control condition, in which participants were not asked to retrieve memories at the time of item rating. These findings suggest that when participants’ attention was drawn to the fact that they could not think of a specific example of themselves behaving in a particular way, they

---

**Table 5A**

Experiment 4 model output: Positive-scored item ratings by retrieval (control condition only).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.468</td>
<td>0.075</td>
<td>-0.060</td>
<td>3.321-3.615</td>
<td>46.354</td>
<td>&lt; 0.001</td>
<td>689</td>
</tr>
<tr>
<td>Memory retrieval</td>
<td>0.319</td>
<td>0.079</td>
<td>0.317</td>
<td>0.164-0.473</td>
<td>4.041</td>
<td>&lt; 0.001</td>
<td>689</td>
</tr>
</tbody>
</table>

Random Effects

- $\sigma^2$: 0.499
- $\tau_{00 pid}$: 0.474
- ICC: 0.487
- N pid: 100
- Observations: 790
- Marginal R²: 0.015
- Conditional R²: 0.495

**Note.** Memory retrieval was dummy-coded (1 = success; 0 = failure).

**Table 5B**

Experiment 4 model output: Reverse-scored item ratings by retrieval (control condition only).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.853</td>
<td>0.076</td>
<td>0.033</td>
<td>3.704-4.001</td>
<td>50.919</td>
<td>&lt; 0.001</td>
<td>679</td>
</tr>
<tr>
<td>Memory retrieval</td>
<td>-0.427</td>
<td>0.111</td>
<td>-0.406</td>
<td>-0.646 to -0.209</td>
<td>-3.839</td>
<td>&lt; 0.001</td>
<td>679</td>
</tr>
</tbody>
</table>

Random Effects

- $\sigma^2$: 0.587
- $\tau_{00 pid}$: 0.486
- ICC: 0.453
- N pid: 100
- Observations: 780
- Marginal R²: 0.014
- Conditional R²: 0.461

**Note.** Memory retrieval was dummy-coded (1 = success; 0 = failure).

**Table 6A**

Experiment 4 model output: Positive-scored item ratings (successful retrieval in retrieve condition vs. all items in control condition).

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Item rating</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.527</td>
<td>0.067</td>
<td>-0.148</td>
<td>3.395-3.659</td>
<td>52.551</td>
<td>&lt; 0.001</td>
<td>1209</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>0.289</td>
<td>0.097</td>
<td>0.308</td>
<td>0.098-0.479</td>
<td>2.990</td>
<td>0.003</td>
<td>197</td>
<td></td>
</tr>
</tbody>
</table>

Random Effects

- $\sigma^2$: 0.464
- $\tau_{00 pid}$: 0.391
- ICC: 0.458
- N pid: 199
- Observations: 1408
- Marginal R²: 0.023
- Conditional R²: 0.470

**Note.** Experimental condition was dummy-coded (1 = retrieve; 0 = control).
inferred that they must not behave that way very often. Under “normal” circumstances, however (i.e., the control condition), the absence of a specific example does not have as large of an impact on how one evaluates oneself on a trait measure such as the TEQ. This conclusion is interesting with respect to another study on memory retrieval and ratings on trait adjectives (Schwarz et al., 1991). In that study, participants were asked to retrieve either six or twelve examples of themselves acting in accordance with a certain trait (e.g., assertiveness), and then had to rate themselves on that trait. Those who recalled twelve memories gave themselves lower ratings than those who recalled six; recalling twelve memories is more difficult than recalling six, and it was argued that the subjective experience of having difficulty finding examples of oneself acting in a certain way led people to think they must not embody that trait as much (Schwarz et al., 1991). With the caveat that the findings of Schwarz et al. (1991) were based on trait adjectives and ours were based on concrete behaviours, the present experiment supports the notion that it is not only the content of memories but the ease of their retrieval that informs self-knowledge (Schwarz & Strack, 2016; see also Tversky & Kahneman, 1973).

We also found that among successfully retrieved memories, those that were more vivid had a stronger influence on item ratings; this finding may be related to greater confidence in the veracity of one’s memories (e.g., Rubin et al., 2003; Talarico & Rubin, 2007).

Table 6B

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.813</td>
<td>0.071</td>
<td>0.181</td>
<td>3.674–3.951</td>
<td>54.044</td>
<td>&lt; 0.001</td>
<td>977</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.578</td>
<td>0.109</td>
<td>-0.561</td>
<td>-0.794 to -0.363</td>
<td>-5.305</td>
<td>&lt; 0.001</td>
<td>184</td>
</tr>
</tbody>
</table>

Random Effects

| σ²               | 0.563     |
| τ00 pid          | 0.425     |
| ICC              | 0.430     |
| N pid            | 186       |
| Observations     | 1163      |
| Marginal R²      | 0.070     |
| Conditional R²   | 0.470     |

Note. Experimental condition was dummy-coded (1 = retrieve; 0 = control).

Table 7A

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.940</td>
<td>0.149</td>
<td>-0.012</td>
<td>2.646–3.233</td>
<td>19.666</td>
<td>&lt; 0.001</td>
<td>603</td>
</tr>
<tr>
<td>Vividness</td>
<td>0.266</td>
<td>0.041</td>
<td>0.218</td>
<td>0.185–0.347</td>
<td>6.432</td>
<td>&lt; 0.001</td>
<td>603</td>
</tr>
</tbody>
</table>

Random Effects

| σ²               | 0.403     |
| τ00 pid          | 0.209     |
| ICC              | 0.342     |
| N pid            | 158       |
| Observations     | 762       |
| Marginal R²      | 0.048     |
| Conditional R²   | 0.373     |

Table 7B

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>SE</th>
<th>std. Beta</th>
<th>CI</th>
<th>t</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.702</td>
<td>0.185</td>
<td>-0.005</td>
<td>3.338–4.066</td>
<td>20.001</td>
<td>&lt; 0.001</td>
<td>329</td>
</tr>
<tr>
<td>Vividness</td>
<td>-0.153</td>
<td>0.055</td>
<td>-0.131</td>
<td>-0.261 to -0.045</td>
<td>-2.778</td>
<td>0.006</td>
<td>329</td>
</tr>
</tbody>
</table>

Random Effects

| σ²               | 0.504     |
| τ00 pid          | 0.317     |
| ICC              | 0.386     |
| N pid            | 124       |
| Observations     | 454       |
| Marginal R²      | 0.016     |
| Conditional R²   | 0.396     |

overall, the findings from Experiment 4 replicate the relationship between episodic memory retrieval and TEQ item ratings observed in Experiments 1–3 (in young and neurologically healthy adults) and extend those studies by showing that this relationship holds when retrieval is confirmed at the time of rating instead of afterward. Our findings also show that the vividness of memories, as well as the failure to retrieve a memory, can influence such judgments.
deficits in episodic autobiographical memory and MTL function show that there is a continuum of self-knowledge, with more context-specific (Never) to 5 (Always). Vividness scale ranges from 1 (Least vivid) to 4 (Most vivid).

Note scoring direction.

Leading theorists have argued that episodic memory is important for keeping track of personal goals rather than self-knowledge per se, with the latter depending on semantic memory instead (Conway, 2005; Conway & Pleydell-Pearce, 2000; Klein & Lax, 2010). Much of the work supporting this notion examines how people judge themselves in relation to abstract trait adjectives (for a review see Klein et al., 2008). The present studies, however, focused on more specific, concrete behaviours that express a trait (items on the Toronto Empathy Questionnaire; TEQ), and in that context we found that episodic memory can inform self-knowledge. It is conceivable that deciding whether we identify with a general trait (e.g., “I am empathic”) is less dependent on recalling specific scenarios compared to judging how frequently we exhibit particular behaviours (e.g., “I can tell when others are sad even when they do not say anything”, an item on the TEQ). Our findings align with the proposal that there is a continuum of self-knowledge, with more context-specific forms of it more closely related to episodic memory and MTL function than others (e.g., Renoult et al., 2012; see also Preble et al., 2013).

More broadly, our findings align with evidence that people with deficits in episodic autobiographical memory and MTL function show abnormalities in their sense of identity. For example, one study found that people with Alzheimer’s disease produced fewer and vaguer statements on an “I Am” task compared to healthy controls, and this was related to their autobiographical memory impairments (Addis & Tippett, 2004). Another study found that people with MTL epilepsy showed an impoverished and restricted sense of identity on the Ego Identity Process Questionnaire (Ballisteri, Busch-Rossnagel, & Geisinger, 1995), again related to deficits in autobiographical memory (Allebone et al., 2015). Interestingly, episodic autobiographical memory contributes to the imagination of specific future scenarios, which is also impaired in people with MTL damage (for reviews see Mullally & Maguire, 2014; Schacter et al., 2012), and answering questions about one’s identity may involve imagining possible versions of oneself (e.g., Dunkel & Anthis, 2001), perhaps more so than does judging how frequently one displays particular behaviours (as in the present studies). The studies by Addis and Tippett (2004) and Allebone et al. (2015) did not link specific memories to individual items the way we did in the present studies, and it is interesting to consider that the effects they found may be mediated by deficits in future imagination. Future research should explore this possibility.

Our findings also align with evidence that older adults, while not showing an impoverished sense of identity, rely less on episodic autobiographical memory to support their identity than do young people (Rice & Pasupathi, 2010; Wolf & Zimprich, 2015). This age-related difference may occur because older adults have had more experiences, enabling them to abstract general knowledge about themselves from many individual episodes (Moscovitch et al., 2005), and may consider such general knowledge to be more informative. In addition, older adults are no longer in the stage of identity formation in which young adults find themselves (Cramer, 2017; Erikson, 1956; Kroger et al., 2010). It has been argued that episodic memory is important for goal pursuit (Conway, 2005; Conway & Pleydell-Pearce, 2000), and if identity formation is not a current goal in the lives of older adults, any episodic memories spontaneously cued may be deemed irrelevant to the task at hand. Either or both of these phenomena may explain why, while we found that older adults did not report significantly fewer episodic memories compared to young adults, their memories did not influence their trait empathy judgments. Relatedly, we found the sensory detail of young adults’ memories to be predictive of their self-reported trait empathy, and while this relationship was numerically weaker in older adults, it was not significantly weaker. The same results were obtained in Experiment 3 for healthy controls and mTLE patients, respectively. The role of the phenomenological quality of memories will be explored below.

Across all of the present experiments, with the exception of Experiment 1, we found that either the sensory detail (Experiment 2), scene coherence (Experiment 3), or vividness (Experiment 4) of participants’ memories was related to self-reported trait empathy. The relationship with vividness in Experiment 4 is likely the most robust finding, given the much larger sample size. In previous work, sensory detail and scene coherence have shown some variability in how strongly they are associated with empathy-related measures (e.g., Campbell et al., 2021; Vollberg et al., 2021). Research on the phenomenology of episodic memories has considered sensory detail and scene coherence (or location clarity) to be aspects of overall vividness, along with other contextual information (e.g., Arnold et al., 2011; D’Argembeau & Van der Linden, 2006, 2012; Szpunar & McDermott, 2008). Others have suggested that vividness ratings may index the salience or availability of particular details, as opposed to the total amount or number of them (Cooper, Kensinger, & Ritchey, 2019; D’Angiuli et al., 2013; Sawczak et al., 2019; Thakral et al., 2020). Overall vividness, therefore, may be a more reliable predictor of self-report ratings compared to measures such as sensory detail and scene coherence, even if it is a less precise construct. Future work would do well to address these questions more closely.

With that said, one possible explanation for the link between vividness and trait empathy judgments is that people tend to be more confident in the veracity of their memories when they are more vivid (regardless of how accurate they really are; Rubin et al., 2003; Talarico & Rubin, 2007). Our participants, therefore, may have considered more vivid memories to be more reliable evidence that they frequently exhibit the empathic behaviours in question. More vivid memories also tend to be more emotional (Talarico, Labar, & Rubin, 2004; Talarico & Rubin, 2003), and our participants may have considered more emotional memories to be more representative of their behaviour. The vividness,
emotionality, and personal significance of autobiographical memories are closely related and are all associated with neural activity in the MTL (Addis et al., 2004; Addis & Schacter, 2008; Gilboa et al., 2004; Moscovich et al., 2005). One of the limitations of the present studies is that we did not measure emotionality or personal significance and were unable, therefore, to examine whether those qualities contribute to trait empathy judgments independent of vividness. Relatedly, spontaneous memories tend to be more positive emotionally (Bernsten, 1998) and less central to one's life story (Cole, Staugaard, & Berntsen, 2016) than voluntary ones. Measuring those constructs in future work comparing different types of retrieval may be informative.

Another limitation is that, given our focus on memories for specific, individual events, we did not account for repeated or generalized event memories (e.g., Holland, Addis, & Kensingten, 2011). For example, the TEQ item, “It upsetts me to see someone being treated disrespectfully”, might remind someone that they have witnessed a friend being belittled by their partner on many occasions. While they may not be able to isolate one particular occasion on which it occurred, and may even retrieve a conglomerate of images from different episodes, their memory of that general experience may influence their rating on that item. Interestingly, a neuroimaging study found that while memories for both specific and repeated events activated an area of the prefrontal cortex associated with self-referential processing, specific memories led to stronger activation (Levine et al., 2004). With respect to the MTL, however, another neuroimaging study found that activation scaled with the amount of detail in autobiographical memories, irrespective of whether they were unique or repeated events (Addis et al., 2004). Nevertheless, as mentioned above, self-knowledge comprises a spectrum ranging from unique episodes to repeated experiences to general facts about oneself (Renoult et al., 2012), and future research should take these gradations into account, along with the type of self-judgment being made (e.g., regarding abstract traits vs. concrete behaviours). It should also be noted that autobiographical memory is supported not only by the MTL but by a network of brain regions, including the medial prefrontal cortex (Andrews-Hanna, Smallwood, & Spreng, 2014; Svoboda, McKinnon, & Levine, 2006), which has been linked to self-referential thinking (Craik et al., 1999; Gusnard, Akhdudak, Shulman, & Raichle, 2001). Neuroimaging work could elucidate how the MTL interacts with the medial prefrontal cortex (and other regions) in the relationship between memory and self-knowledge.

The present studies show that retrieval of episodic autobiographical memories can amplify judgments about related aspects of one’s own personality in healthy young adults with intact MTLs. Related past work from Schwarz et al. (1991), however, found that when participants were explicitly required to retrieve 12 as opposed to six memories, their self-report judgments were dampened, even when they were successful in retrieving more memories. This finding suggests, in addition to the memory being accessed, the ease of retrieval also informs such judgments (Schwarz & Strack, 2016). This interpretation is broadly consistent with our finding in Experiment 4 that if retrieval failed when it was explicitly required, ratings were dampened even more than when it was not required. Considering these findings together suggests the possibility of an inverted U-shape relationship between directed episodic memory retrieval and self-report judgments: Perhaps retrieving one or a few episodic memories amplifies self-report judgments, compared to retrieving none at all (at least for healthy young adults), but as more episodic memories are expected, and retrieval becomes more difficult, there is a dampening effect. Critically, both Schwarz et al. (1991) compared the directed retrieval of six and 12 memories, they did not compare six to fewer or to zero memories, nor did they explicitly distinguish between episodic and semantic memories. It would be interesting for future research to compare various numbers of episodic memories to test an inverted U-shape model of the relationship between episodic memory and self-report judgments.

In sum, the present work points to a reliable relationship between episodic autobiographical memory retrieval and self-reported trait empathy on a questionnaire. This relationship is stronger when memories are more vivid, and is diminished in older adults and individuals with mTLEs. In contrast to prominent work in the field, our findings suggest that episodic memory is linked to self-knowledge, at least when people are reflecting on specific behaviours that express a trait, as opposed to general trait concepts. A natural next step is to use the present approach to investigate whether episodic memory can influence self-report judgments of other traits in addition to empathy. Preliminary evidence from our laboratory on a version of the Big Five personality inventory suggests that it does. Future research should also take a wider survey of the phenomenological qualities of memories and account for different levels of event specificity. It may also be useful to include a measure of identity development status, such as the Ego Identity Process Questionnaire (Balistreri et al., 1995), as individuals who are wrestling the most with defining their identities may show greater reliance on episodic memory in making self-report judgments. Most of the models in the present paper have fairly low marginal $R^2$ values, and accounting for these other factors may increase the amount of variance explained. Finally, it should be noted that the present findings do not go so far as to suggest that episodic memory influences actual empathic behaviour via changes in self-knowledge. However, when the present studies are viewed together with previous research, in which episodic processes affect prosocial behaviour such as economic donations to people in need (Gaesser, Keeler, & Young, 2018, Exp. 3; Gaesser, Shimura, & Cicarn, 2020; Exp. 5), it is worth considering this possibility. Future work could use informant reports (e.g., Vazire, 2010) or other measures (such as charitable donations) to explore potential effects on actual behaviour. The relationship between memory and the self has long been recognized as a complex and important one, and we look forward to future studies that will illuminate it further.

Competing interests statement

None of the authors have any conflicts of interest to declare.

Data statement

Data will be available at [https://doi.org/10.6084/m9.figshare.13370204](https://doi.org/10.6084/m9.figshare.13370204) upon publication.

CRediT authorship contribution statement

Caspian Sawczak: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. Mary Pat McAndrews: Resources, Writing – review & editing, Supervision, Funding acquisition. Brendan Bo O’Connor: Conceptualization, Writing – review & editing, Supervision, Funding acquisition. Zoe Fowler: Software, Investigation, Data curation, Writing – original draft. Morris Moscovich: Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This research was supported by grants from the Canadian Institutes of Health Research (M.M., #8347) and the Natural Sciences and Engineering Research Council (M.P.M., RGPIN-2015-06471), a Templeton Science of Prospection Award (B.G.), and University of Toronto graduate student fellowship (C.S.). We would also like to thank Irene Gianoulis for assistance with recruiting mTLE patients at Toronto Western Hospital, Jennifer Stellar for helpful comments on an early version of this manuscript that appeared in a doctoral dissertation (Sawczak, 2020), and Elisa Giaramelli, Marliney Ziegler, Astrid Coleman, and members of the Moscovich and McAndrews labs at the University of Toronto, for helpful discussion of this work.