Sensorimotor Effects in Surprise Word Memory – a Registered Report

Stage 2 Submission

Agata Dymarska[1,2], Louise Connell[3]

[1] Faculty of English, Adam Mickiewicz University, Poland

[2] Department of Psychology, Lancaster University, UK

[3] Department of Psychology, Maynooth University, Ireland

Author Note

Agata Dymarska https://orcid.org/0000-0003-0931-3081

Louise Connell https://orcid.org/0000-0002-5291-5267

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682848) to LC.
Correspondence concerning this article should be addressed to Agata Dymarska, Faculty of English, Adam Mickiewicz University, Poznań, Poland, or to Louise Connell, Department of Psychology, Maynooth University, Co. Kildare, Ireland. Email: agata.dymarska@amu.edu.pl or louise.connell@mu.ie

URL to the registered Stage 1 manuscript: https://osf.io/ck5bg

Abstract

Sensorimotor grounding of semantic information elicits inconsistent effects on word memory, depending on which type of experience is involved, with some aspects of sensorimotor information facilitating memory performance while others inhibit it. In particular, information relating to the body appears to impair word recognition memory by increasing false alarms, which may be due either to an adaptive advantage for survival-relevant information (whereby words pertaining to the body spread activation to other concepts and generate a confusable memory trace) or to a somatic attentional mechanism (whereby words pertaining to the body activate a false sense of touch that renders their representations less distinctive as memory trace and retrieval cue). To date, the existing literature does not distinguish between these two explanations. We set out to adjudicate between them using a surprise memory task, where participants study the words under a guise of a lexical decision task, which allowed us to examine how participants form a memory trace for words grounded in bodily experience. We found support for the somatic attentional account, as body related words increased false alarms even when attention was not directed to them at the study phase. Overall, the results provide further evidence for the importance of distinctiveness in word memory trace, and suggest a reinterpretation of the role of semantic richness in word memory.

Keywords: word memory; sensorimotor information; semantic richness; incidental memory

Introduction

Memory for words is facilitated by semantic content of their representations. As proposed by semantic richness theory (Buchanan et al., 2001; Pexman et al., 2008), a richer representation, such as a larger number of features or associates, higher body-object interaction ratings, or higher imageability, makes it easier to process and respond to the word (e.g., Cortese & Fugett, 2004; Dunabeitia et al., 2008; Pexman et al., 2003; Siakaluk et al., 2008). This idea has been extended to memory research, where a richer representation makes the memory trace more distinctive: that is, it increases the amount of conceptual knowledge associated with an item that is not shared with other items in the list (Dobbins & Kroll, 2005). Specifically, a word richer in semantic information elicits stronger semantic activation (Pexman et al, 2013), and therefore produces a stronger memory trace (Hargreaves et al., 2012; Sidhu & Pexman, 2016). Such rich, distinctive representations are more likely to be retained until the test phase (Lau et al., 2018) and correctly recognised as old, because their memory trace will not easily fade or be replaced by interfering information. Indeed, it has been demonstrated that word memory is better for items with higher imageability (Cortese et al., 2010; 2015), body-object interaction (Sidhu & Pexman, 2016), higher animacy and perceived threat (Bonin et al., 2014; Leding, 2020), and stronger sensorimotor grounding involving experience with manipulable objects and food concepts (Dymarska et al., 2023a). Additionally, more distinctive and semantically rich words are also less likely to be mistaken for previously-seen words (see e.g., Glanzer & Adams, 1985; Zechmeister, 1972), leading to lower false alarms for words higher in imageability or arousal (Cortese et al., 2010; 2015; Lau et al., 2018; but cf. Ballot et al., 2021, regarding imageability). Such studies show that semantically-rich, distinctive words tend to facilitate recognition memory in the classic mirror pattern (Glanzer & Adams, 1985) of increasing hit rates and reducing false alarms.

Notably, semantic information facilitates memory above and beyond lexical word characteristics, which can facilitate memory performance in their own right when they make the word more distinctive. Repeatedly encountered-high frequency words, or words with high contextual diversity (i.e., encountered in multiple contexts) tend to be remembered better (Cortese et al., 2015) as their distinctive trace is less likely to be confused with other items. Similarly, words which have fewer orthographic or phonological neighbours are less confusable and easier to remember (Cortese et al, 2004; 2015; Glanc & Greene, 2007). Nonetheless, word meaning – that is, semantic representation of the word's referent – is activated automatically upon encountering a word, and its effects on memory performance have been found even when lexical characteristics such as word frequency have been taken into account (Dymarska et al., 2023a; Glanzer & Adams, 1985; Sidhu & Pexman, 2016).

However, when it comes to the perception and action experience that underlies representations of word meaning, there is a wide range of sensorimotor dimensions that can provide grounding (Lynott et al., 2020), and these various forms of sensorimotor experience do not all elicit straightforward semantic richness effects on word memory. In Dymarska et al. (2023a), we compressed ratings of sensorimotor strength in 11 dimensions (i.e., measuring the extent to which a concept is experienced with each of 6 perceptual modalities and 5 action effectors: Lynott et al., 2020) into four orthogonal (uncorrelated) principal components and investigated their effect on recognition memory for over 5000 words (data from Cortese et al., 2010; 2015). Although all forms of sensorimotor experience were initially expected to make a word's representation more distinctive and hence facilitate recognition memory performance, as per other semantic richness variables, we instead found that results were mixed. Words that scored highly on Object experience (e.g., *pillow, comb*; involving vision, touch, and hand/arm action) or Food experience (i.e., *pastry, omelette*; involving taste, smell, and mouth action) were indeed remembered better. Consistent with semantic richness effects, higher scores in these sensorimotor components resulted in more hits, fewer false alarms, and overall improved recognition memory performance (i.e., higher HR-FA and *d'*). However, Communication experience (e.g., *chat, rumor*; involving hearing, interoception, head action and mouth action) had no discernible effect on recognition memory, as higher scores in this component made no difference to either hits or false alarms, which Dymarska et al. (2023a) suggest may be due to lack of distinctiveness in Communication-related words. Most strikingly, words that scored strongly on the Body component (e.g., *cuddle, fitness*; involving touch, interoception, and action of the hand/arm, foot/leg, and torso) *impaired* rather than facilitated memory performance. Instead of producing the mirror pattern predicted by semantic richness theory of increased hits and fewer false alarms, higher Body scores unexpectedly had little effect on hits but led to *more* false alarms, overall worsening recognition memory (i.e., negative HR-FA and *d'*). That is, semantic richness from the Body component actually led participants to mistakenly respond to new words as if they had been previously seen.

One possible explanation for such divergent sensorimotor effects is that the unexpected effect of the Body component resulted from an adaptive advantage for survival-relevant information. It has been suggested that stimuli pertaining to our survival, such as threatening sounds, wild animals, or bodily sensations (e.g., *movement* or *pain*, which scored highly on the Body component) automatically and preferentially capture attention and spread activation to other, interconnected concepts that may increase the chance of survival, which offers an adaptive advantage in threat detection from an evolutionary perspective (Howe & Derbish, 2010; Nairne et al., 2007; 2008). While this process leads to a strong memory trace for studied words (thereby increasing hit rates), the activation of other, related concepts has the side effect of increasing the probability of false recall and inflating false alarms in recognition memory (Howe & Derbish, 2010; Leding, 2019; 2020; Otgaar & Smeets, 2010). Since bodily function and integrity is, by definition, important to survival, this account offers a potential reason why Dymarska et al. (2023a) found elevated false alarms for Body-related words. That is, even though the semantic representation of a strongly Body-related word might be distinctive by itself, its tendency to activate networks of other survival-relevant concepts lowers its distinctiveness and leads to a more confusable memory trace and less-effective retrieval cue, hence inflating false alarms when new Body-related words in the test phase are mistakenly matched to those granted a memory trace via spreading activation.

However, finding this adaptive effect for Body-related words in a standard recognition memory task, which does not specifically require attending to survival-related information, may be an artefact of the experimental paradigm used. Previous observations of inflated false alarms for survival-related words come from experiments that place particular emphasis on contextual elaboration during the study phase, such as by asking participants to rate the usefulness of presented words for survival in a grasslands scenario (e.g., Howe & Derbish, 2010; Otgaar & Smeets, 2010), or to generate ideas associated with the presented words (Bonin et al., 2022). Although participants in Dymarska et al. (2023a) and in Leding (2020) were not instructed to elaborate on the meaning of the study words, they knew that they were going to be tested later on their memory for the words, and so were likely to employ whatever strategies they could to remember the words better, such as elaborating their meaning and mentally imagining them in a context that would be easy to remember (e.g., a threatening scenario). Hence, when learning words that involve bodily experience in the study phase, such strategies could have led to activation of related concepts in the memory trace and inadvertently reduced participants' ability to correctly discriminate between the survival-relevant words they studied and new (survival-relevant) stimuli presented at the test phase. Body-related words may therefore inflate false alarms and impair memory performance only when learned with contextual elaboration, meaning that this pattern of effects would be unlikely to occur within a memory paradigm where participants were not motivated to elaborate on a word's representation during the study phase.

Alternatively, it is possible that the inflated false alarms in Dymarska et al. (2023a) were due to a somatic attentional mechanism which modulates perception and representation of body-related experiences. Directing attention to the hand (Mirams et al., 2010), to interoceptive sensations (i.e., heartbeat: Mirams et al., 2012), or to locations within peripersonal space (Mirams et al., 2013) all caused people to mistakenly believe they were perceiving tactile sensations (i.e., a micro pulse being delivered to their finger) when none were present. In other words, directing attention towards the body in various ways leads people to false alarm on touch sensations. Several other perceptual phenomena have been found to reappear in semantic processing of word stimuli (e.g., modality switching costs: Pecher et al., 2003; tactile disadvantage in stimulus detection: Connell & Lynott, 2010), and grounded cognition theories hold that that the conceptual system has co-opted the sensorimotor system for the purposes of representation (e.g., Barsalou, 1999; Connell & Lynott, 2014). Hence, the somatic attentional account also offers a potential reason why Dymarska et al. (2023a) found elevated false alarms for Body-related words. That is, if attending to the body (or space close to the body) in one perceptual modality activates a false sense of touch, then the presence of hand/arm action and interoceptive experience in Body-related words may similarly cause additional, irrelevant tactile activation when presented in a recognition memory task, rendering the overall representation of the word less distinctive and prone to false alarms. Even though the semantic representation of a strongly Body-related word might be distinctive in its own right, its tendency to activate other modalities of body-related experience (particularly touch) lowers its distinctiveness and leads to a more-confusable memory trace and a less-effective retrieval cue. Unlike the adaptive account, the attentional account does not depend on the experimental paradigm used to investigate word recognition memory. If Body-related words inflate false alarms and impair memory performance because they tend to activate other modalities of bodily experience such as touch (i.e., creating a less-distinctive memory trace), then it will occur regardless of how participants are instructed in the study phase.

Dymarska et al. (2023a) could not distinguish between these possibilities due to analysing a conventional expected memory task, where participants deliberately learned a list of words during the study phase because they expected to be tested on them later. When participants learn a list of words with the explicit knowledge their memory will later be tested, they are likely to employ strategies to help them remember the words such as semantic and contextual elaboration. Conversely, when participants are *not* aware they will be later tested on their memory for presented words (i.e., a surprise memory task), such elaboration is far less likely, and offers us an opportunity to adjudicate between theoretical accounts and adapt the semantic richness theory to account for this additional mechanism. Table 1 illustrates the mechanisms behind the two theoretical accounts, and their predictions are outlined in detail in the Current Study section.

Table 1. Mechanisms underlying the two proposed accounts of the Body effects given an expected and a surprise memory task.

Theoretical account	Expected task: study phase = deliberate word learning (Dymarska et al., 2023a)	Surprise task: study phase = lexical decision word/nonword (the present RR)
Adaptive advantage	Word learning leads to strategic elaboration of word meaning, which triggers spreading activation for survival-relevant words → Body-	Lexical decision leads to minimal elaboration of meaning, and so minimal spreading activation for survival-relevant words → Body-

	related words lose distinctiveness in memory trace and inflate FA	related words remain distinctive in memory trace and either do not inflate FA or actively lower FA.
Somatic attention	Reading words automatically grounds meaning in sensorimotor information, which then activates irrelevant modalities for words concerned with bodily states → Body-related words lose distinctiveness in memory trace and inflate FA	Reading words automatically grounds meaning in sensorimotor information, which then activates irrelevant modalities for words concerned with bodily states → Body-related words lose distinctiveness in memory trace and inflate FA

Current Study

In the current study, we are going to test the effects of multiple aspects of sensorimotor experience on a surprise (or incidental) word recognition memory task, which will allow us to distinguish the adaptive and the attentional accounts of the Body effects on word memory via contrasting predictions. As in Cortese et al. (2010; 2015), we take a megastudy approach (Balota et al., 2012) to word recognition memory by testing an item set of over 5000 words. The advantage of this approach is that it allows for a large-scale analysis using multiple, continuous predictors, without the need to match word samples on each of the characteristics in different factorial conditions. As predictors of word recognition memory, we employ the same lexical and sensorimotor variables analysed in Dymarska et al. (2023a) that were obtained via Principal Components Analysis of a large number of lexical variables with multidimensional sensorimotor dimensions.

To create a surprise memory paradigm, we will use a lexical decision task to present words to participants in the study phase (i.e., without informing them they will later be tested on their memory for these words), and then conduct the test phase using a regular old/new recognition task. Our reason for using a lexical decision task (i.e., decide whether or not a string of letters is a valid word) is that the semantic activation of different aspects of sensorimotor information has recently been established (Dymarska et al., 2023b): sensorimotor components reflecting Body, Communication, Objects and Food experience all facilitate lexical decision to varying extents. Based on these findings, we expect that semantic information from all four sensorimotor components will be activated in the study phase of our surprise memory task and thereby contribute to the memory trace of each word, but their influence on the recognition memory decision in the test phase will vary by component.

In line with semantic richness theory, we predict that encountering words whose referents involve object- and food-related experience will activate a rich semantic representation, which will support a distinctive memory trace in the study phase and serve as an effective retrieval cue in the test phase. That is, we expect the current surprise memory task to produce the same effects as in Dymarska et al. (2023a)'s expected memory task, where words scoring highly on Object and Food components will be remembered better and will be discriminated easily between old (studied) and new words (i.e., higher hit rates and lower false alarms). However, based on previous findings for word recognition memory in Dymarska et al. (2023a), we expect that – despite its activation during lexical decision (Dymarska et al., 2023b) – words involving communication experience will not produce particularly distinctive representations, meaning such words are neither very memorable nor provide informative cues for correct discrimination between studied and non-studied words. As a result, the Communication component will not elicit any effects on word memory performance.

Critically, when it comes to words relating to body experience, we aim to disentangle the underlying reasons for their inflated false alarms observed in Dymarska et al. (2023a). In an expected memory task, participants are likely to use elaboration as a strategy to make the word memorable, for example by placing a concept in a particular scenario or context. According to the adaptive advantage explanation, this process will trigger survival-relevant words to spread activation to a network of other, related concepts, which will make the memory trace for those words less distinctive and prone to false alarms. However, participants are unlikely to use elaboration as a strategy in a surprise memory task because they do not know that they will be tested on memory for the words. Therefore, without such elaboration, spreading activation to related concepts will not take place for survival-relevant words, and they will no longer be prone to false alarms. It is possible that the pattern of the Body component effects in Dymarska et al.'s (2023a) expected task was due to such an elaboration strategy, but if so, the same pattern will not emerge in the surprise memory task.

By contrast, the somatic attention account does not rely on elaboration at encoding and so is unaffected by the task manipulation. In both expected and surprise memory tasks, when word meaning is automatically accessed on reading during the study phase, any representations relating to bodily states will extend attention to touch and other irrelevant modalities, which will make the memory trace for those words less distinctive and prone to false alarms. It is possible that the pattern of the Body component effects in Dymarska et al.'s (2023a) expected task was due to such automatic processes, and if so, the same pattern of will emerge in the surprise memory task.

Hypotheses

Our specific hypotheses concern the following dependent measures of word recognition memory: hit rate (HR), false alarm rate (FA), overall performance (HR-FA), and sensitivity (*d'*):

1. Higher scores on the Food and Object components will facilitate memory performance, leading to better recognition of old words (positive effect on HR), better rejection of new words (negative effect on FA), and better discrimination of old versus new items (positive effect on HR-FA and *d'*).
2. Higher scores on the Communication component will not affect performance in any variable.
3. Finally, higher scores on the Body component will produce different predictions by theory. If the somatic attention account applies to word memory, there will be a positive effect on both HR and FA, and a negative effect on discrimination (HR-FA and *d'*). However, if the adaptative account is correct, then there will be a positive effect on HR only and either no effect (or negative effect, in a mirror pattern) on FA, that will potentially be strong enough to produce a positive effect on discrimination (HR-FA and *d'*).

Method

Participants

We determined sample size using sequential hypothesis testing with Bayes Factors (BF: e.g., Schönbrodt et al., 2017), where BF is computed repeatedly during data collection until it exceeds an *a priori* defined threshold of evidence or the maximum feasible sample size is reached. We begun analysing data at Nmin = 20 participants per list (total Nmin = 2120), which is the sample size in Cortese and colleagues' dataset (for an expected recognition memory task) that allowed us to detect the key Body effect on FA (Dymarska et al., 2023a). Our maximum sample was 60 participants per word (6360 participants in total); this sample size, which was times larger than the Nmin, was chosen to allow us to detect effect sizes even if they are smaller in the surprise memory task than in the expected memory task. The critical threshold of evidence was set to BF ≥ 6 (or its reciprocal BF≤ 1/6). Sequential sampling proceeded in increments of 106 participants to maintain a balanced design across stimulus lists, and was due to stop when the BF evidence for or against the inclusion of each sensorimotor component in the model clears the critical threshold for all four DVs[1], or until we reach the maximum sample. In other words, the stopping rule in our sequential hypothesis testing plan required all sensorimotor predictors for all DVs to be simultaneously out of the equivocal zone in order to stop testing. In case at any point one predictor dropped below the threshold of $BF_{10} = 6$ (and remained above the reciprocal $BF_{01} = 1/6$), even if it previously cleared it at a smaller sample, we planned to continue testing until we found stable evidence for or against all effects, which would provide a robust estimate of the true effect of each sensorimotor component on memory for words, or until we reach Nmax. In the event that a particular effect was still not detected at Nmax, we planned to conclude that it is likely too small to be of interest.

Our stopping rule was triggered at a sample size of $N = 23$ participants per list (total final sample $N = 2438$). Native speakers of English without a reading impairment (i.e., dyslexia) were recruited via Prolific.co platform. We recruited participants with prior approval rate on Prolific of at least 95%. Participants were excluded from analysis according to criteria outlined in the Data Analysis section. Participants were paid £1.40 for their time (i.e., approximately £8.50/hour pro-rata).

Ethics and Consent

The study received ethical approval from the Lancaster University Faculty of Science and Technology Research Ethics Committee (reference number FST18088). Before taking part in the study, participants were asked for their informed consent to participate. They first read information detailing the purpose and expectations of the study, which described the task as being about "word judgement" (i.e., without mentioning the surprise memory test). Consent included agreement that payment is on condition of passing two attention checks (i.e., a screen during the distractor task and

---

[1] The Stage 1 registered report stated five DVs due to an error in editing; the correct number is four because the hypotheses concern only four measures of recognition memory performance (i.e., HR, FA, HR-FA, *d'*).

after the last trial of the test phase that asks participants to "press the letter Q/P as quickly as possible"), and following the instructions (i.e., participants would be excluded for pressing the same key throughout a task, e.g., giving all word or all nonword responses in the lexical decision task; all correct or incorrect responses to the equations in the distractor task; all old or all new responses in the recognition memory task), and that all data will be shared publicly in anonymised form. All materials, anonymised data, analysis code and full results are available on the Open Science Framework at https://osf.io/hqd2m/.

Materials

We took a sample of 5300 words for which sensorimotor effects were previously analysed by Dymarska et al. (2023a) in an expected recognition memory task using data from Cortese et al. (2010; 2015). In this way, we could use the same words and the same predictor variables as Dymarska et al. (2023a), so any changes in sensorimotor effects can be accorded to task differences rather than item differences. Dymarska et al. analysed 5305 words in total; we dropped five words with the lowest UK prevalence score (Brysbaert et al., 2019) to produce a final set of 5300 words that could be divided into 106 equal lists of 50 target words each. In order to ensure that similar words are distributed across different lists, we used a binned sampling method when dividing the stimuli into word lists so that all lists contained items that spanned from low to high scores on all 6 components. We sampled from bins (set as component score quartiles) of each orthogonal component, such that every list included 3 words from each quartile of each component (i.e., 3 words x 4 quartiles x 4 components = 48 words), plus 2 words selected at random from different bins to bring the total to 50 words per list.

For the lexical decision task in the study phase, we also generated 5300 corresponding pseudowords using Wuggy (Keuleers & Brysbaert, 2010). Wuggy produces pronounceable nonwords that follow the phonotactic constraints of English and match the length and number of syllables of the target words. The pseudowords are produced by changing one or more phonemes in real target words (e.g., "church" → "chulks"). We divided these pseudowords into 106 lists of 50 pseudowords each, where each pseudoword list corresponded to (i.e., was derived from) one of the 106 target word lists. In order to ensure a given target word was presented separately from its derived pseudoword, and was different from the subsequent distractor list (see below) each target word list was paired with a non-corresponding pseudoword list. For example, target word list 1 was paired with pseudoword list 3, target word list 2 was paired with pseudoword list 4, target word list 3 was paired with pseudoword list 5, and so on. This process produced 106 list pairs (each containing 50 words and 50 pseudowords) for presentation in the lexical decision task, where each participant will see a single list pair.

For the surprise memory task in the test phase, each target word list was paired with another target word list which served as distractors (i.e., "new" words not seen in the study phase), and vice versa. That is, target word list 1 was presented with word list 2 as distractor list, and served as distractor list for target list 2. Target word list 3 was presented with word list 4 as distractors, and served as distractor list for target list 4, and so on. This process produced 106 list pairs (each containing 50 "old" target words and 50 "new" distractor words) for presentation in the recognition memory task.

For a distractor phase (i.e., in between the study and test phases), we created a list of simple mathematical problems, as per Cortese et al., (2010; 2015). These problems comprised 18 simple addition and subtraction equations for verification (e.g., "2 + 3 = 6?"), all using single digit numbers, where half of the equations were correct and half were incorrect.

Finally, for the statistical analysis, we used the same six predictor variables used by Dymarska et al. (2023a) to analyse sensorimotor effects on word recognition memory. These variables comprised component scores derived from a principal components analysis (PCA) to consolidate a large set of intercorrelated lexical and sensorimotor variables into orthogonal (i.e., uncorrelated) predictors. The components were originally obtained in a previous study (Dymarska et al., 2023b); full details can be found there, here we summarise the method of extracting the components. The item set for the PCA was based on 9796 words used in the analysis of imageability on visual word recognition by Dymarska et al. (2023b). Variables included in the PCA can be seen in Table 2, and covered a wide range of lexical and semantic characteristics of the words. Critically, they incorporated 11 dimensions of sensorimotor strength from the Lancaster Sensorimotor Norms (Lynott et al., 2020), where each dimension comprised a rating of the extent to which the word's referent was experienced with the specified perceptual modality or by performing an action with the specified action effector. In addition, they also included Lynott et al.'s composite measure of all 11 dimensions, Minkowski-3 sensorimotor strength, which was weighted towards the dominant dimension(s) for a given word. PCA (parallel analysis at 95th percentile, correlation matrix, varimax rotation) reduced the original 24 dimensions to an optimal 6 orthogonal components that captured 77.4% of the original variance. Each component was labelled according to the variables that loaded upon it, producing two components that represented lexical characteristics of the word (Frequency and Length) and four components that represented sensorimotor experience with the referent concept (Body, Object, Food, Communication). Table 2 summarises how each component relates to the original variables which were entered into the PCA. These six components cleanly distinguished between lexical and semantic information, with the exception of the Object component, which included the noun (part of speech) variable in addition to sensorimotor variables; this contribution was not unexpected since object concepts are typically labelled with nouns (e.g., *apple*, *dog*) and tend to be strongly experienced with visual, haptic, and hand/arm action. Table 3 shows examples of words scoring highest and lowest on each component.

Table 2: Variables used by Dymarska et al. (2023b) in Principal Components Analysis and the rotated components (used here as predictors) to which they most strongly contributed with positive or negative weighting ($r > .3$ or $< -.3$).

Original variable	Source	Definition	PCA component
LgSUBTLWF	ELP	Log word frequency (US English)	+Frequency
LgSUBTLCD	ELP	Log contextual diversity (how many contexts a word appears in; US English)	+Frequency
Zipf Frequency	Van Heuven et al. (2014)	Word frequency on Zipf scale (UK English)	+Frequency
Prevalence	Brysbaert et al. (2018)	How many people know the word (probit value)	+Frequency
Familiarity	Stadthagen-Gonzales & Davis (2006); Scott et al. (2018); Wilson (1988)	How subjectively familiar a word seems (ratings)	+Frequency
Age of Acquisition	Kuperman et al. (2012)[a]	Approximate age that the word was learned	–Frequency
Linguistic distributional distance (LDD20)	Dymarska et al., 2023b	Distributional neighbourhood (mean cosine distance to closest 20 neighbours, based on vectors of log co-occurrence frequency)	–Frequency
Word length	ELP	Word length in letters	+Length
Number of syllables	ELP	Word length in syllables	+Length
Orthographic Levenshtein Distance (OLD20)	ELP	Orthographic neighbourhood (mean letter Levenshtein distance to closest 20 neighbours)	+Length
Phonological Levenshtein Distance (PLD20)	ELP	Phonological neighbourhood (mean phoneme Levenshtein distance to closest 20 neighbours)	+Length
Torso action strength	LSN	Motor strength in torso effector	+Body
Foot/leg action strength	LSN	Motor strength in foot/leg effector	+Body
Hand/arm action strength	LSN	Motor strength in hand/arm effector	+Body, +Object
Composite sensorimotor strength	LSN	Aggregated sensorimotor strength in all dimensions (Minkowski-3 distance of 11-dimension vector from the origin)	+Body, +Object, +Communication, +Food
Head action strength	LSN	Motor strength in head effector	+Communication
Auditory strength	LSN	Perceptual strength in hearing modality	+Communication
Mouth action strength	LSN	Motor strength in mouth effector	+Communication, +Food
Gustatory strength	LSN	Perceptual strength in taste modality	+Food
Olfactory strength	LSN	Perceptual strength in smell modality	+Food
Visual strength	LSN	Perceptual strength in sight modality	+Object
Noun (part of speech)	ELP	Whether or not word is a noun (binary coded: noun=1, non-noun=0)	+Object
Haptic strength	LSN	Perceptual strength in touch modality	+Object, +Body, –Communication
Interoceptive strength	LSN	Perceptual strength in interoceptive (sensations inside the body) modality	–Object, +Body, +Communication

[a] With extended norms from http://crr.ugent.be/archives/806

Note: ELP = English Lexicon project (Balota et al., 2007); LSN = Lancaster Sensorimotor Norms (Lynott et al., 2020).

Table 3: Top five (highest scoring) and bottom five (lowest scoring) words for each component.

Component	High scoring	Lowest scoring
Frequency	the, that, and, what, about	slat, adage, welt, jeer, vise
Length	friendship, transplant, somewhere, privilege, threshold	rap, sang, pun, gab, hum
Body	move, movement, bathe, strength, pain	because, about, but, than the
Food	meal, pizza, pastry, omelette, pasta	waltz, listen, chase, polka, ballet
Object	nail, dog, pillow, pistol, cat	quench, queasy, hungry, nauseous, digest
Communication	song, concert, joke, word, chat	dorsal, fertile, which, than, enzyme

Procedure

The experiment was created and hosted through the online experiment builder Gorilla (http://www.gorilla.sc/; Anwyl-Irvine et al., 2019). Following informed consent and demographic questions about first language, age, sex and handedness, it consisted of three stages: a study phase, a distractor task, and a test phase.

In the study phase, participants were asked to perform a lexical decision task, where they had to decide whether a presented string of letters was a real word in English (e.g., “coat”) or not (e.g., “soat”) by pressing the “Z” key (not a real word) or “M” key (real word) on their keyboard. Each trial began with a blank screen for 200 ms, followed by a fixation cross for 300 ms, and then the word (or pseudoword) presented individually in lowercase in the centre of the screen, 14 pixels in size, using black text on a white background, in Open Sans font (see Figure 1). The (pseudo)word stayed onscreen until the participant responded, or until a timeout limit of 3000 ms was reached. A short practice task with 4 items was presented before the main task, where participants received feedback for accuracy and speed; if they did not respond within 3000 ms, the message “Too slow” was displayed for 1000 ms and the next trial commenced. If they responded on time, a green tick (for correct responses) or a red cross (for incorrect responses) was displayed for 1000 ms before the next trial started. There was no feedback during the main lexical decision task. The order of (pseudo)words was randomised for each participant.

Once the study phase (lexical decision task) was over, participants were presented with a distractor task and asked to verify 18 mathematical equations, as per Cortese et al., (2010; 2015). Participants proceeded through questions by pressing the “Z” key (incorrect) or “M” key (correct) to respond at their own pace, without feedback. The entire distractor task took approximately 20-30 seconds.

Finally, in the test phase following the distractor task, participants performed a surprise recognition memory task. Participants were asked to decide whether or not they saw each displayed word earlier in the study (i.e., in the lexical decision task) by pressing the “Z” key (new word) or “M” key (old word) on their keyboard. Words were displayed in randomised order as per the lexical decision task (see Figure 1) and stayed onscreen until the participant responded or until a timeout limit of 3000 ms was reached. There was no feedback during the recognition memory task. We measured accuracy of responses and response times from the onset of each word until participant pressed a response key; accuracy per word was used to calculate the dependent measures of memory performance (see Data Analysis Plan) and RT was used for data exclusions. At the end of the test phase, participants saw a short debriefing screen that thanked them for taking part and they were then redirected to the recruitment platform. The mean length of the entire procedure from consent to debrief was 7 minutes 39 seconds. This mean duration includes loading time for the experiment, which was delayed for some participants due to capacity issues on the Gorilla experimental platform. Running of the experiment was unaffected by delays in loading.

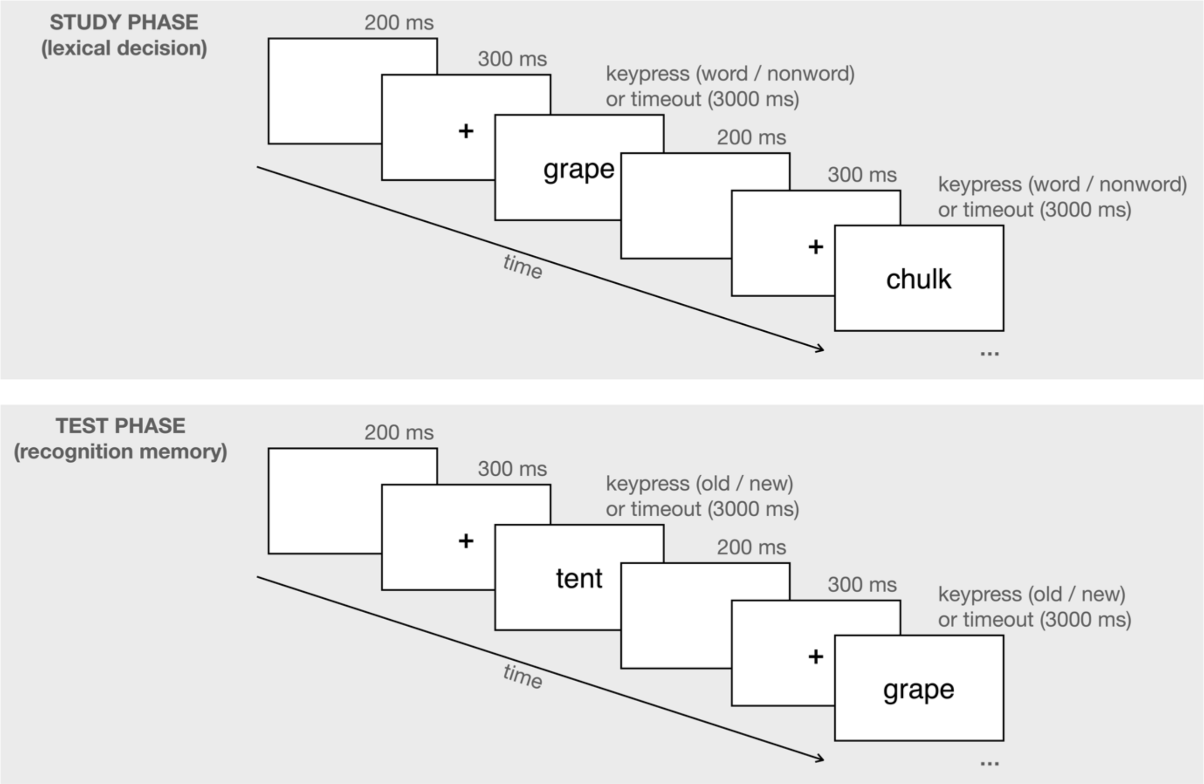

Figure 1: Trial sequence diagrams for the lexical decision task in the study phase (upper panel) and the recognition memory task in the test phase (lower panel).

Data Analysis Plan

Data Exclusions

We replaced any participants who were rejected without payment (see Ethics and Consent section, N=57) and participants who timed out on more than 30% of trials in the study or test phases (N=8). We also replaced participants who did not reach 60% overall accuracy (i.e., across hit rates and correct rejections) on the recognition memory task, in line with Cortese et al. (2010; 2015) (N=659). This accuracy threshold took into account both old words where participants correctly respond “old” (i.e., hit rates to targets) *and* new words where participants correctly respond “new” (i.e., correct rejections of distractors = 1 – false alarms), and thus subsumed a threshold based on *d’*. We expected that any participants who did not perform well on the lexical decision task, for example due to inattention, would also not perform well on the memory task, and therefore we did not plan to use separate exclusion criteria based on lexical decision task accuracy.

Trials with RT below 300ms were classified as motor errors and removed (763 responses; 0.31% of data). Trials which did not register a response within 3000ms timed out and were removed from the analysis (1225 responses; 0.50% of data). No other data were excluded as outliers, and no outlier correction method was performed.

Dependent Variables

We calculated the following memory performance measures per item, as in Cortese et al. (2010; 2015): hit rate (HR: proportion of times studied word was correctly classified as "old"), false alarms (FA: proportion of times new word was incorrectly classified as "old"), hit rate minus false alarms (HR-FA: overall performance per word), and sensitivity (*d'*: sensitivity per word). The signal detection variable *d'* measures the sensitivity in discriminating a given word as old versus new, and (unlike HR-FA) is unaffected by underlying response bias; it was calculated using a log-linear approach (Stanislaw & Todorov, 1999) to compensate for floor/ceiling performance (i.e., HR or FA at 0% or 100%).

Statistical Analysis

We analysed memory performance per item by running hierarchical linear regressions on each dependent variable. The hierarchical model structure allowed us to compare our results with previously-found effects of sensorimotor components on expected word memory performance (Dymarska et al. 2023a). In Step 1, we entered Frequency and Length components as baseline predictors. In Step 2 we added the four sensorimotor components (Body, Communication, Food, Object). There were four regressions in total, one for each measure of memory performance (HR, FA, HR-FA, *d'*).

Bayesian linear regressions were conducted in JASP (*version 0.18.01*: JASP Team, 2022), with default JZS priors (r = .354) and a Bernoulli distribution (p = 0.5), from which we report BFs for model comparisons between hierarchical steps and inclusion BFs of coefficients (i.e., relative likelihood of models including a particular predictor compared to models excluding it). The threshold for inference was $\text{BF}_{\text{inclusion}} \geq 6$ for evidence *in favour* of a particular component's effect, or its reciprocal $\text{BF}_{\text{inclusion}} \leq 1/6$ for evidence *against* its effect. Where evidence was in favour, the direction of the relevant coefficient determined whether a given sensorimotor component elicits a facilitatory (or inhibitory) effect on word memory performance.

In order to perform an exploratory comparison of the effect sizes per component in the current study with the results of Dymarska et al. (2023a), we conducted frequentist linear regressions in JASP with the same model structure as above and extracted part correlations per component per dependent variable. We report these part correlations to illustrate the unique contribution each predictor makes to the dependent measure in question.

[STAGE 2 - NEW TEXT STARTS HERE]

Results

Overall, performance on the memory task was good, with high hit rates and low false alarms, similar to the expected memory task analysed in Dymarska et al. (2023a) (see Table 4 for descriptive statistics).

Table 4: Mean performance on each memory measure in Study 1 with its standard deviation.

DV	Mean	SD
Hit rate	0.769	0.136
False alarms	0.176	0.128
HR-FA	0.593	0.187
d'	1.558	0.820

Lexical effects at Step 1 were largely consistent with previous research. Lower frequency words produced higher hit rates and HR-FA and better *d'* sensitivity (as in Cortese et al., 2010; 2015; Higham et al., 2009; Lau et al., 2018). Lower frequency also led to lower false alarms, similar to Glanzer and Adams (1985) and Lau et al. (2018), but unlike the expected memory task in Dymarska et al., (2023a; see also Cortese et al., 2010, 2015), suggesting that low-frequency words tended to create a more distinctive memory trace. Word length produced small effects on word memory performance, but the pattern of results indicated that shorter words elicited lower HR and FA, and lower *d'* discrimination, with no effect on HR-FA. Full statistics are available in supplemental materials on the OSF.

Overall, the Step 2 model with sensorimotor components offered a much better fit for each of the DV compared to the baseline model with Frequency and Length only ($BF_{10} = 2.73 \times 10^{16}$; see Table 5).

Table 5: Variance in memory performance explained by each step of the regression models (change in R^2, with levels of Bayesian evidence), and uniquely explained by each sensorimotor component in the Step 2 model (squared part correlations).

Model / parameter	HR	FA	HR-FA	*d'*
Step 1: Lexical baseline R^2	0.350***	0.004**	0.206***	0.302***
Step 2: Sensorimotor ΔR^2	0.013***	0.028***	0.037***	0.014***
Body	0.004	0.002	0.006	0.004
Communication	0.001	0.011	0.010	0.003
Food	0.006	0.013	0.017	0.005
Objects	0.002	0.001	0.003	0.002
Total R^2	0.363	0.032	0.243	0.316

* $BF_{10} > 6$, positive evidence; ** $BF_{10} \geq 20$, strong evidence; *** $BF_{10} \geq 150$, very strong evidence

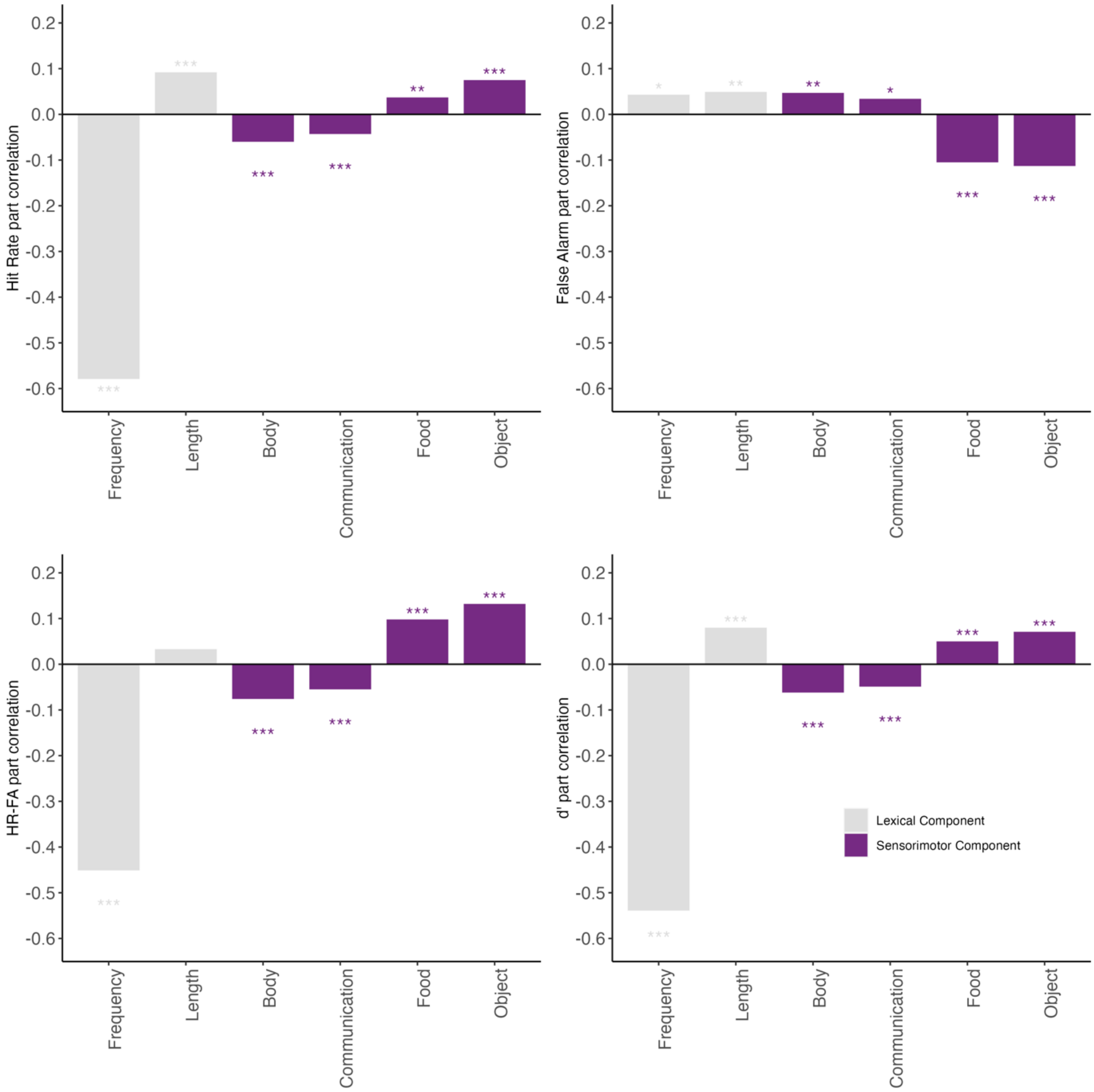

Figure 2: Part Correlations of Each Component Predictor. Asterisks indicate the inclusion Bayes factor (BF) of each predictor: *** $BF_{inclusion} \geq 150$, constituting very strong evidence; ** $BF_{inclusion} \geq 20$, strong evidence; * $BF_{inclusion} > 6$, positive evidence. HR-FA = hit rate minus false alarm rate.

Hypothesis 1: Food and Object Components

The effects of the Food and Object components on word memory performance were as predicted: both facilitated memory performance. That is, words scoring highly on the Food and Object components were better recognised and discriminated, as indicated by positive effects on HR, HR-FA and *d'*, and were more likely to be correctly rejected when they were new words, as indicated by a negative effect on FA (all $BF_{inclusion} > 40.6$). These results were consistent with semantic richness theory predictions, as in Dymarska et al. (2023a), whereby sensorimotor experience relating to taste, smell, and mouth action (Food component), as well as vision, touch and hand/arm action (Object component), enhanced the memorability of words.

Hypothesis 2: Communication Component

Contrary to expectations, the Communication component elicited strong effects on memory performance measures, *impairing* memory performance in each case (see Figure 2). Words scoring highly on the Communication component were less likely to be correctly recognised when old (i.e., negative effect on HR), more likely to elicit false alarms when new (i.e., positive effect on FA), and overall, were poorly discriminated (i.e., negative effects on HR-FA and *d'*); all $BF_{inclusion} > 6.94$. In other words, participants exhibited worse memory for words relating to Communication experience (i.e., sound and interoceptive experience, as well as mouth and head action). These effects were contrary to our expectations of null effects, as previously reported for an expected memory task (Dymarska et al. 2023a), but nonetheless consistent with high-Communication words having low distinctiveness; we return to this point in the General Discussion.

Hypothesis 3: Body Component

The effects of the Body component largely follow the pattern predicted by the somatic attention account. There was a critical *positive* effect on FA (i.e., inflating false alarms; $BF_{inclusion} = 106.59$), which resulted in a negative effect on discrimination measures (HR-FA and *d'*; all $BF_{inclusion} > 4.80 \times 10^5$). However, higher Body scores led to a negative effect on HR (i.e., poorer correct recognition of old words; $BF_{inclusion} = 3.67 \times 10^5$), rather than the positive effect originally predicted. Nonetheless, since we predicted a positive effect on HR regardless of theoretical account, due to previous findings for an expected memory task (Dymarska et al., 2023a), this unanticipated negative effect does not differentiate between the somatic attention and adaptive accounts. Results are consistent with the idea that processing word meaning related to bodily states leads to attention being extended to tactile and other somatically-relevant modalities, which makes its representation less distinctive as a memory trace and retrieval cue, and therefore prone to false alarms. We discuss this account in light of unexpected results in the General Discussion.

Exploratory Analysis

We ran two exploratory analyses: The first one was an analysis of RT to explore how the sensorimotor components affected the time taken to correctly recognise a previously-studied word. Since RT analysis is rarely included in word recognition memory studies, we had no specific predictions about how sensorimotor strength would influence response times. Nonetheless, we wanted to make the RT results available for researchers who are interested in investigating the timecourse of word recognition memory processes. The second analysis was a trial-level Bayesian regression analysis of responses made to explore how the sensorimotor components affected the sensitivity of words being judged as old versus new. This approach provides parameter estimates for lexical and sensorimotor effects on *d*', while accounting for participant-level and item-level variability. While this analysis does not directly address the question of how the Body component affects the false alarm rate – critical to distinguish competing theoretical accounts of word recognition memory – it offers additional insight into the robustness of lexico-semantic effects on discrimination performance.

We first conducted an exploratory analysis of RT for correct "old" responses (i.e., hits). We standardized RT via z-score transformation per participant and then calculated the mean zRT per word in order to minimize participant variability in response latencies (see e.g., Balota et al., 2007). The resulting zRT was then analysed in an item-level Bayesian linear hierarchical regression with the same steps and predictor variables as the confirmatory analysis. The data favoured the Step 2 model with sensorimotor components over the Step 1 lexical model ($BF_{10} = 5.740\times10^{30}$), where both the Food ($BF_{inclusion} = 6.542\times10^{11}$) and Object components ($BF_{inclusion} = 3.715\times10^{19}$) facilitated RT of correctly recognising a previously-seen word. There was equivocal evidence for the effects of Body and Communication components on RT ($BF_{inclusion}$ = 1.941 and 0.355, respectively). Full coefficient statistics are in supplemental materials.

We next employed Bayesian mixed-effect probit regression using the brms package (Bürkner, 2021) to analyse the response made on each trial (i.e., 0 = new, 1 = old), with random effects of participants and items and fixed effects of word type (0 = new, 1 = old), the two lexical and four sensorimotor components, and the interaction of all components with word type. In such a model, the effect of word type represents estimates of *d'* and interactions of a given component with word type estimate effects of that component on *d'* (DeCarlo, 1998; Rouder & Lu, 2005). Within that model, we found that lower Frequency increased the likelihood of correct discrimination (i.e., greater sensitivity in correctly distinguishing old versus new words), while greater Length increased it slightly. For the sensorimotor components, Body and Communication components *decreased* the likelihood of the words being correctly distinguished, while the Food and Object component increased it. These results are in line with the item-level confirmatory analysis of *d'*; we include full details in supplemental materials.

General Discussion

We set out to analyse the effects of sensorimotor information on a surprise task of word recognition memory and determine why information relating to the body impairs word memory performance. Previous research indicated that while some aspects of sensorimotor experience (i.e., relating to food and objects) exerted typical semantic richness effects, facilitating word memory by creating a strong, distinctive memory trace, other aspects did not (Dymarska et al., 2023a). In particular, experience relating to the body overall impaired memory performance by inflating false alarms. The present registered report tested the same large set of words in a surprise memory paradigm, where participants studied the words incidentally in a lexical decision task rather than deliberately with foreknowledge of a memory test. Results showed that body experience continued to increase false alarms and overall impair recognition memory performance, which supports the hypothesis that such effects are due to a somatic attentional mechanism that leads body-related words to activate tactile and other bodily modalities that renders their representations less distinctive as a memory trace and retrieval cue.

Specifically, we found that, similar to previous research, semantic richness elicited varying effects on memory for words, depending on the sensorimotor experience underlying their meaning. The Food and Object components facilitated memory by creating a strong and distinctive memory trace, in line with the predictions of semantic richness theory (Buchanan et al., 2001; Pexman et al., 2008). Critically, the effects of the Body component shed light on the question of somatic attention versus adaptive accounts and their role in memory for words related to bodily action, interoception, and touch. As found in Dymarska et al. (2023a), words scoring highly on the Body component impaired recognition memory by increasing the number of false alarms, which explicitly supports the somatic attention account. According to this account, when participants read body-related words, it directs attention towards their bodily experiences as part of representing their meaning, which in turn spreads activation to modalities and effectors related to somatic experience beyond what is relevant for the current word, in particular tactile sensations. In the study phase, when participants encounter body-related words during lexical decision, this spread of somatic attention leads to a less distinctive memory trace. When encountered in the test phase, it makes the displayed word less effective as a retrieval cue because it leads participants to associate the displayed word with a broader sensorimotor trace, creating a feeling of familiarity. As a result, participants tend to mistake new words for old (i.e., increasing the rate of false alarms) and overall discriminate poorly between old and new items (i.e., decreasing HR-FA and *d*'). In this way, semantic richness effects on memory are constrained by distinctiveness (Dymarska et al., 2023b; Lau et al., 2018), whereby semantically rich, distinctive representations facilitate recognition memory but semantically rich, confusable representations are apt to impair memory performance due to imprecise overlap between the cue and the memory trace.

However, not all hypotheses were fully supported. First, we originally hypothesised that the Body component would have a positive effect on hit rates because that is the pattern that emerged in previous work using an expected memory task (Dymarska et al., 2023a). Instead, we found a negative effect: that is, words scoring higher on the Body component were less likely to be correctly recognised in the test phase. Although the direction of this effect was unexpected, it is still consistent with the somatic attention account in that the spread of somatic attention leads body-related words to have a less distinctive memory trace, and lower distinctiveness in the memory trace often leads to lower hit rates. For instance, common words tend to be less distinctive than rare words, which is why lexical frequency variables often have negative effects on hit rates (e.g., the present study; Lau et al., 2018). We acknowledge that this explanation for the Body effect on hit rates is somewhat post hoc, but it is important to note that neither theoretical account predicted a negative effect a priori. The

Body effect on false alarms in the present study is what allows us to adjudicate between the somatic attention account (supported) and the adaptive account (not supported). If the Body component had failed to inflate false alarms in this surprise memory task, where participants are not motivated to elaborate on word meanings to make them more memorable, then evidence would have favoured the adaptive account instead. The fact that a clear positive effect emerged for false alarms is what leads us to conclude that the present findings support the somatic attention account. In light of this broader pattern of effects, the negative effect on hit rates for body-related words is not inconsistent with the somatic attention account.

The second set of unanticipated effects are those of the Communication component, where previous research led us to hypothesise no effects on word recognition memory, but we instead found that higher Communication scores impaired memory, decreasing hit rates and increasing false alarms. In the expected memory task, the Communication component elicited no effects on any variable, and Dymarska et al. (2023a) speculated that the lack of effects stemmed from lack of distinctiveness. That is, due to high-Communication words tending to cluster around similar meanings (e.g., *joke, pun*; *song, music*), they do not produce particularly distinctive representations and so they are neither particularly memorable nor informative when attempting to discriminate between old and new words. Notably, it is also possible that high-Communication words are drawing on somatic attention effects, albeit to a lesser extent than Body words. Somatic experience is particularly associated with interoception (i.e., sensations inside the body), which loads positively on the Body component but also on the Communication component at broadly similar strength (Dymarska et al., 2023b). Other forms of somatic experience load positively on either component without loading negatively on the other; for instance, action of the torso, hand/arm, foot/leg and haptic experience load on Body while action of head and mouth load positively on Communication. That is, somatically relevant concepts will score highly on the Body component if concerned with whole-body movement and/or interception (e.g., *bathe, cuddle, strong, cold*) but some of them, depending on their meaning, will potentially instead score highly on the Communication component if concerned with head-mouth movement and/or interoception (e.g., *song, cry, sneeze, laugh*). For this reason, it is possible that somatic attention leads to irrelevant somatic activation for high-Communication words as well as for high-Body words, hence making their representations less distinctive and producing similar patterns of effects. Since this explanation is based on unexpected rather than hypothesised effects, it is best regarded as speculative. Nonetheless, the unexpected effects in the present study – where high Communication scores impaired word memory – are consistent with this idea of low distinctiveness. Similar patterns were reported by Lau et al. (2018), who found that more highly polysemous words (i.e., with greater number of senses) have lower rates of free recall and poorer discriminability in recognition memory, which they proposed was because their ambiguous representations are less distinctive and lead to more confusable memory traces. Semantically rich words are only effective as retrieval cues when their representations are also distinctive (e.g., for Food and Object components) but when not distinctive, the rich representations are prone to lose their effectiveness (e.g., the Communication component in Dymarska et al.'s expected memory task) or even actively impair performance due to mistaken overlaps with similar memory traces (e.g., the Communication component in the present surprise memory task; the Body component).

Nonetheless, the findings raise the question: why do sensorimotor effects differ between the present surprise memory task and the previous expected memory task (Dymarska et al., 2023a)? Memory performance was broadly similar in the two tasks, which would suggest that removing participants' opportunity to deliberately learn lists of words does not necessarily impair their memory for those words. However, the baseline lexical model explained markedly more variance in performance in the present surprise paradigm (e.g., 35.0% of HR variance; 30.2% of *d'*) than in the previous expected paradigm (e.g., 26.1% of HR; 8.2% of *d'*). This pattern of effects is consistent with a greater focus on lexical characteristics of words in the encoding phase of the present study, as one might expect in a lexical decision task that requires participants to distinguish words from non-words. Conversely, the Step 2 sensorimotor model explained *less* variance in the surprise paradigm (e.g., 1.3% of HR; 1.4% of *d'*) than the previous expected paradigm (e.g., 4.7% of HR; 5.9% of *d'*). It is important to note that differences in sensorimotor effects between the two tasks are not due to methodological artefacts. For instance, in the present surprise task we analysed the same set of words as Cortese and colleagues' (2010, 2015) original data collection for the expected task, in the same design and list size, with comparable sample sizes per list, using the same experimental procedure with the exception of our critical task change in the encoding phase from deliberate learning to lexical decision. Moreover, the statistical analysis and predictor variables we used in the present paper for the surprise memory task were identical to those used to analyse the expected memory task in Dymarska et al. (2023b). We therefore interpret this drop in sensorimotor explanatory power as reflecting the relatively shallow semantic processing of words during encoding in the surprise paradigm (i.e., lexical decision) compared to the deeper elaborative processing in the expected paradigm (i.e., deliberate study). That is, people focused more on the wordforms and less on the meaning in a surprise memory task compared to when they were learning words for a later memory test.

In terms of individual sensorimotor components, the Food and Object components generally followed the overall pattern of sensorimotor effects, facilitating performance in the surprise paradigm with smaller effect sizes than in the expected paradigm. Notably, the two components where distinctiveness appears to be an issue – Body and Communication – showed the reverse pattern, with numerically *larger* effects in the present surprise task than in the expected memory task, and of course both impairing performance rather than facilitating it. We speculate that the effortful elaboration of semantic representations during encoding in the expected paradigm may have compensated for the lack of distinctiveness of words relating to Body and Communication. By making their representations slightly more distinctive (or at least less indistinctive) during encoding, it is possible that the expected paradigm effectively suppressed the extent to which the Body and Communication components impaired recognition memory. That is, it may be the case that depth of processing affects the extent to which semantic richness is constrained by distinctiveness, whereby low distinctiveness could be more detrimental to memory performance in shallow semantic processing than in deep semantic processing. Future research should examine this possibility in more detail.

Overall, the results support the reinterpretation of sematic richness theory in memory, where stronger semantic richness does not always lead to facilitation of memory due to the complex, multistep nature of the memory encoding, retrieval and recognition processes, and the multidimensional nature of semantic representations which are remembered differently depending on the form of sensorimotor experience that underlies word meaning. Additionally, taking together the results of the expected memory task and the surprise memory task, we conclude that the effects of the Body component can be explained by the somatic attention account. That is, regardless of the nature of the task, word meaning is automatically accessed when seeing the word, and the activation (when relating to the body) spreads to touch and related (but irrelevant) modalities, reducing the distinctiveness of the memory trace for those words and leading to an uninformative cue, resulting in poor discrimination and poor recognition performance. Future research on word memory should focus on other semantic variables which have been found to facilitate lexical processing, in order to determine whether and how the effects extend to memory, with a view to deepening understanding of how semantic richness influences memory processes.

References

Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods, 52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ballot, C., Mathey, S., & Robert, C. (2021). Word imageability and orthographic neighbourhood effects on memory: a study in free recall and recognition. *Memory*, *29*(6), 829-834. https://doi.org/10.1080/09658211.2021.1921216

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Nelson, D. L., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459. https://doi.org/10.3758/bf03193014

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing?. In: Adelman JS (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology.* Hove, England: Psychology Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bonin, P., Gelin, M. & Bugaiska, A. (2014). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Memory & Cognition*, *42*, 370–382. https://doi.org/10.3758/s13421-013-0368-8

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*, 467-479. https://doi.org/10.3758/s13428-018-1077-9

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin and Review*, *8*(3), 531–544. https://doi.org/10.3758/BF03196189

Bürkner, P. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software, 100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Connell, L., & Lynott, D. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*(1), 1-9. https://doi.org/10.1016/j.cognition.2009.10.005

Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*, *143*(2), 527–533. https://doi.org/10.1037/a0034626

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 384–387. https://doi.org/10.3758/BF03195585

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Cortese, M. J., Watson, J. M., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory & Cognition, 32*, 632-639. https://doi.org/10.3758/BF03195854

Dobbins, I. G., & Kroll, N. E. A. (2005). Distinctiveness and the Recognition Mirror Effect: Evidence for an Item-Based Criterion Placement Heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1186–1198. https://doi.org/10.1037/0278-7393.31.6.1186

Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1072-1077. https://doi.org/10.3758/PBR.15.6.1072

Dymarska, A., Connell, L., & Banks, B. (2023a). More is not necessarily better: How different aspects of sensorimotor experience affect recognition memory for words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(10), 1572–1587. https://doi.org/10.1037/xlm0001265

Dymarska, A., Connell, L., & Banks, B. (2023b). Weaker than you might imagine: Determining imageability effects on word recognition. *Journal of Memory and Language, 129*, 104398. https://doi.org/10.1016/j.jml.2022.104398

Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory & Cognition, 35*(2), 365–371. https://doi.org/10.3758/BF03193457

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20. https://doi.org/10.3758/BF03198438

Hargreaves, I. S., Pexman, P. M., Johnson, J. C., & Zdrazilova, L. (2012). Richer concepts are better remembered: number of features effects in free recall. *Frontiers in Human Neuroscience*, *6*, 73. https://doi.org/10.3389/fnhum.2012.00073

Howe, M. L., & Derbish, M. H. (2010). On the susceptibility of adaptive memory to false memory illusions. *Cognition*, *115*(2), 252-267. https://doi.org/10.1016/j.cognition.2009.12.016

JASP Team (2022). *JASP (Version 0.16.x) [Computer software]*.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627-633. https://doi.org/10.3758/BRM.42.3.627

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology*, *71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Leding, J. K. (2019). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*, 383-394. https://doi.org/10.3758/s13421-018-0873-x

Leding, J. K. (2020). Animacy and threat in recognition memory. *Memory & Cognition*, *48*(5), 788-799. https://doi.org/10.3758/s13421-020-01017-5.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods, 52,* 1-21. https://doi.org/10.3758/s13428-019-01316-z

Mirams, L., Poliakoff, E., Brown, R. J., & Lloyd, D. M. (2010). Vision of the body increases interference on the somatic signal detection task. *Experimental Brain Research*, *202*(4), 787-

794. https://doi.org/10.1007/s00221-010-2185-7.

Mirams, L., Poliakoff, E., Brown, R. J., & Lloyd, D. M. (2012). Interoceptive and exteroceptive attention have opposite effects on subsequent somatosensory perceptual decision making. *Quarterly Journal of Experimental Psychology*, *65*(5), 926-938. https://doi.org/10.1080/17470218.2011.636823.

Mirams, L., Poliakoff, E., Brown, R. J., & Lloyd, D. M. (2013). Brief body-scan meditation practice improves somatosensory perceptual decision making. *Consciousness and Cognition*, *22*(1), 348-359. https://doi.org/10.1016/j.concog.2012.07.009

Nairne, J.S., Pandeirada, J.N.S., & Thompson, S.R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science, 19,* 176-180. https://doi.org/10.1111/j.1467-9280.2008.02064.x

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Otgaar, H., & Smeets, T. (2010). Adaptive memory: Survival processing increases both true and false memory in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 1010. https://doi.org/10.1037/a0019402

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*(2), 119-124.

Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*(6), 842-855. https://doi.org/10.3758/BF03196439

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: semantic richness effects in language processing. *Frontiers in Human Neuroscience*, *7*, 723. https://doi.org/10.3389/fnhum.2013.00723

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322-339. https://doi.org/10.1037/met0000061

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2018). The Glasgow norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*(3), 1258-1270. https://doi.org/10.3758/s13428-018-1099-3

Siakaluk, P. D., Newcombe, P. I., Duffels, B., Li, E., Sidhu, D. M., Yap, M. J., & Pexman, P. M. (2016). Effects of emotional experience in lexical decision. *Frontiers in Psychology, 7*, article 1157. https://doi.org/10.3389/fpsyg.2016.01157

Sidhu, D. M., & Pexman, P. M. (2016). Is moving more memorable than proving? Effects of embodiment and imagined enactment on verb memory. *Frontiers in Psychology*, *7*, 1–13. https://doi.org/10.3389/fpsyg.2016.01010

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. https://doi.org/10.3758/BF03207704

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. Behavior Research Methods, *Instruments, & Computers, 20*(1), 6-10.

Zechmeister, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *The American Journal of Psychology, 85*(3), 425-430.

Question	Hypothesis	Sampling Plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes	Observed outcome
What are the effects of multiple aspects of sensorimotor experience on a surprise word recognition task?	1. We hypothesise that sensorimotor experience of food and objects will facilitate memory performance in a surprise memory task. Higher scores on the Food and Object components will lead to higher hit rates, lower false alarms, and better discrimination between old and new items.	We will use sequential hypothesis testing with BF to determine optimal sample size. Nmin is 2120 participants (20 per word list), and we will continue in increments of 100 participants until the evidence for or against the inclusion of all predictors in analyses of all dependent variables clears the BF=6 threshold, or until we reach Nmax of 6360 participants (60 per list). The Nmin sample of 20 participants per word list was sufficient to detect component effects in an expected memory task (see Dymarska et al., preprint). The Nmax represents the largest sample size our resources allow, in case a surprise task requires a larger sample size to detect effects than the expected memory task.	We will conduct item-level hierarchical Bayesian linear regressions on four dependent variables (hit rate, false alarms rate, and two measures of discrimination: HR-FA and *d’*). We will examine the effects of sensorimotor information on word memory performance, above and beyond any effects of lexical information, using six orthogonal components as predictors. Step 1 enters lexical controls of Frequency and Length, and Step 2 enters the critical predictors that represent different types of experience: Body, Communication, Food, and Object.	Hypotheses will be confirmed or disconfirmed according to the inclusion BFs of sensorimotor parameters in the Step 2 model. The BF inferencing threshold of 6 is the required level of evidence allowing to conclude the presence or absence of effects in a registered report submission.	Bayes Factor of 6 or above will indicate that the component elicits an effect on memory performance. Bayes factor below 1/6 will indicate that the component does not influence memory performance (i.e., evidence against the effect). Bayes Factor between these values will indicate equivocal evidence regarding whether or not the component influences memory performance. The direction of the regression coefficient will indicate whether the component enhances memory (positive effect on HR and discrimination measures, negative effect on FA) or impairs memory (negative effects on HR and discrimination measures, positive effects on FA). null Body effect on FA accompanied by null effects of Food	1.If the Food and Objects components elicit null effects (i.e., do not facilitate memory performance), it could suggest that semantic richness effects of food- and object-related experience do not extend to a surprise memory task. In other words, it would suggest that incidental memory and expected memory do not rely on sensorimotor information in the same way. Alternatively, if null Food and Object effects are accompanied by null Body effects, it would suggest regression to the mean relative to the previous expected memory task.	Hypothesis confirmed. Food and Object related experience facilitated word memory performance. Higher scores on the Food and Object components led to higher hit rates, lower false alarms, and better discrimination between old and new items.
	2. We hypothesise that sensorimotor experience of communication will not affect memory performance, that is, high scores on Communication component will not have any effects on hit rate, false alarms or discrimination measures.					2. Any positive evidence for Communication effects would indicate that incidental memory and expected memory do not rely on sensorimotor information the same way. If high scores on the Communication component unexpectedly enhance memory performance, it would counter Dymarska et al.’s (2023a) suggestion that words related to communication experience lack distinctiveness.	Hypothesis not confirmed. Higher Communication experience affected memory for words negatively. Higher Communication scores led to lower hit rates, higher false alarms and poorer discrimination between old and new items.

	3. We hypothesise that higher Body scores will lead to higher hit rates, but the effects that emerge on false alarms and discrimination measures will vary by theory, and will allow us to determine the underlying mechanism.				and Object components indicates RTM	3. If the Body component increases false alarms and decreases HR-FA and *d'*, then the somatic attention explanation is correct and the adaptive advantage explanation is not. Conversely, if the Body component elicits no effect on false alarms or decreases false alarms, and increases HR-FA and *d'*, then the adaptive account applies and the somatic attention explanation is incorrect.	Hypothesis not confirmed. Higher Body scores led to lower hit rates. Additionally, higher Body scores led to higher false alarms and poorer discrimination between old and new items, supporting the somatic attention account of word memory.