Uncertainty aversion predicts the neural expansion of semantic representations

Vives, M.L.; Bruin, D. de; Baar, J.M. van; FeldmanHall, O.; Bhandari, A.

nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01561-5

Uncertainty aversion predicts the neural expansion of semantic representations

Received: 6 May 2022

Accepted: 17 February 2023

Published online: 30 March 2023

Check for updates

Marc-Lluís Vives[1,2] ✉, Daantje de Bruin[1], Jeroen M. van Baar[3], Oriel FeldmanHall[1,4,5] ✉ & Apoorva Bhandari[1,5] ✉

Correctly identifying the meaning of a stimulus requires activating the appropriate semantic representation among many alternatives. One way to reduce this uncertainty is to differentiate semantic representations from each other, thereby expanding the semantic space. Here, in four experiments, we test this semantic-expansion hypothesis, finding that uncertainty-averse individuals exhibit increasingly differentiated and separated semantic representations. This effect is mirrored at the neural level, where uncertainty aversion predicts greater distances between activity patterns in the left inferior frontal gyrus when reading words, and enhanced sensitivity to the semantic ambiguity of these words in the ventromedial prefrontal cortex. Two direct tests of the behavioural consequences of semantic expansion further reveal that uncertainty-averse individuals exhibit reduced semantic interference and poorer generalization. Together, these findings show that the internal structure of our semantic representations acts as an organizing principle to make the world more identifiable.

Human life is rife with uncertainty, from the information we gather through our senses[1,2] to the unpredictable outcomes of our actions[3–6]. People intolerant to uncertainty find it especially aversive and are therefore strongly motivated to reduce it[7,8]. When confronted with ambiguous situations that lack a clear interpretation, those who are averse to uncertainty experience stress and will take action to avoid any additional uncertainty[9]. Because these individuals are more sensitive to uncertainty and perceive greater uncertainty than those who are uncertainty tolerant[10], they are often better at remembering concepts or cues that signal uncertainty[11]. In short, uncertainty aversion plays an outsized role in shaping behaviour across a range of domains[12–15]. Despite this, uncertainty is inescapable; it even imbues the concepts we use to make sense of the world[16]. Take for instance the words 'slip' and 'lapse', which have similar but subtly different meanings. When focusing on the similarities, both refer to some form of mistake, which means that they can be represented similarly in semantic space and it may not matter which word one uses. Focusing on the differences, however, highlights that a 'slip' is usually trivial and accidental, while a 'lapse' can be serious and imply responsibility, and this may cause an individual to separate these concepts in semantic space. In this Article, we propose that individuals seeking to minimize uncertainty distinguish concepts from one another by separating them in semantic space.

Semantic representations encode our conceptual knowledge about the world and influence how we perceive incoming sensory information[17], which enables us to ascribe meaning to stimuli[16]. However, because stimuli typically activate more than one concept, there is always some uncertainty when identifying meaning. The degree of uncertainty between a pair of concepts depends on their separation in semantic representational space. Therefore, ensuring that concepts are sufficiently differentiated from one another in such a space helps to reduce semantic uncertainty. This logic accords with classic work showing that the closer a pair of concepts are in psychological space (for example, snow and cold), the more confusable they are, while more dissimilar concepts are easier to discriminate and thus suffer

[1]Department of Cognitive, Linguistic, Psychological Sciences, Brown University, Providence, RI, USA. [2]Department of Psychology, Leiden University, Leiden, The Netherlands. [3]Trimbos Institute, Netherlands Institute for Mental Health and Addiction, Utrecht, The Netherlands. [4]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI, USA. [5]These authors contributed equally: O. Feldman Hall, A. Bhandari ✉e-mail: m.l.vives.moya@fsw.leidenuniv.nl; oriel.feldmanhall@brown.edu; apoorva_bhandari@brown.edu

https://doi.org/10.1038/s41562-023-01561-5

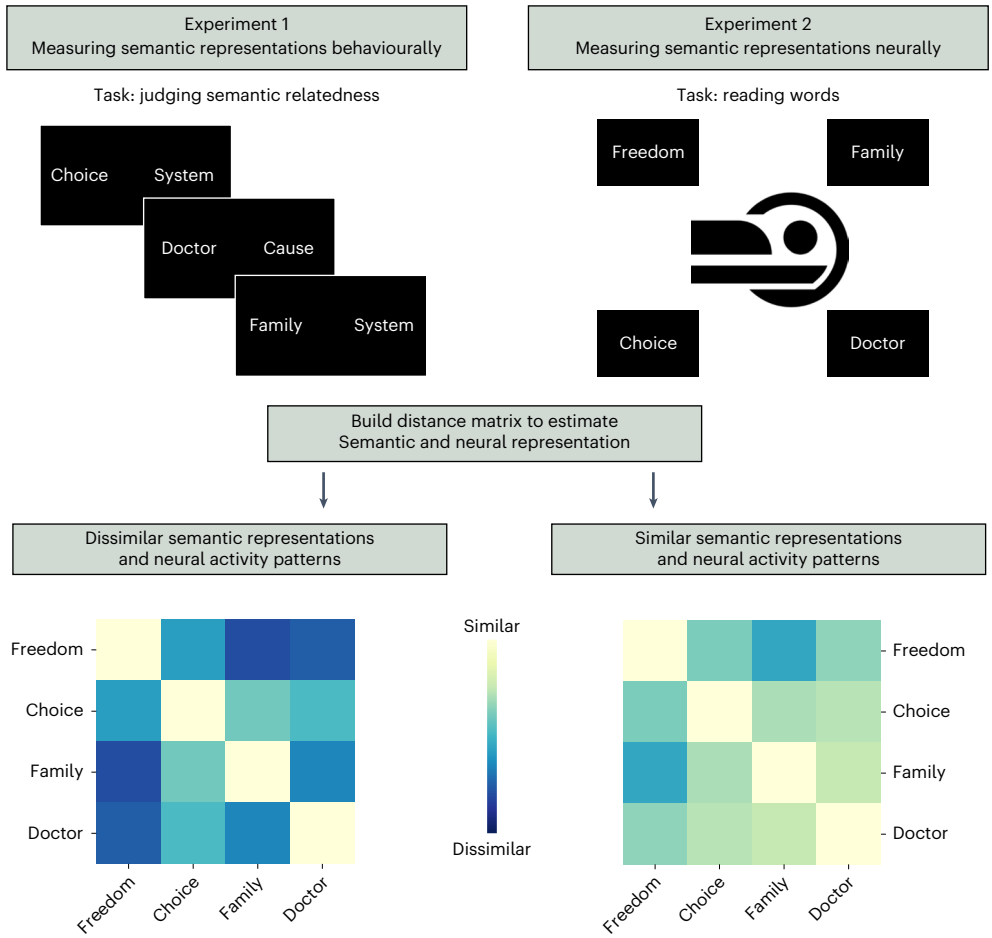

Fig. 1 | Dissimilar semantic representations and neural activity patterns are more identifiable. Each word is associated with a semantic representation captured by judging the semantic relationship between two sets of words (experiment 1) or a unique neural pattern of activity captured by reading a list of words one at a time (experiment 2). Semantic and neural competition increases as similarity between neural activity increases. Neural representations that are more dissimilar on average would suffer less neural competition and, as a consequence, be more identifiable.

far less from semantic interference (for example, cat and black)[18,19]. In short, the distances between concepts in semantic representational space determine whether one's everyday experiences are conceptually ambiguous or fairly clear-cut. We hypothesize that individuals averse to uncertainty mitigate the effects of uncertainty by making their semantic representations more distinct. Here we test whether this process, which we term semantic expansion, causes increases in pair-wise distances between concepts.

The echo of such a strategy should also be detectable in the structure of neural representations. Consider a brain region that encodes concepts in the activity of its neurons. Each concept is associated with a distinct neural activity pattern, and separating the activity patterns from each other as much as possible enables a downstream neuron reading out the representation to disambiguate each concept from the others[20–22]. The greater the distance between neural activity patterns, the more separable the representation is[23,24] and, consequently, the less uncertainty there is in the representation—which would be very useful to those who prefer to reduce the attendant uncertainty of a concept (Fig. 1). Simply put, our semantic-expansion hypothesis predicts that individuals who are averse to uncertainty should exhibit more differentiated neural activity patterns.

While the need to discriminate between concepts encourages concepts to be separated from each other, there is an opposing functional influence on the structure of the representational space: the need to keep similar concepts closer to each other in psychological space. This promotes useful generalization between similar concepts. In other words, the structure of a semantic representation reflects a trade-off between the need for discrimination and generalization. We hypothesize that people averse to uncertainty, through semantic expansion, privilege discrimination over generalization.

To test this semantic-expansion hypothesis, we conducted four experiments in which we examined the effects of individual differences in uncertainty attitudes on semantic distances between words in psychological and neural space and their behavioural consequences. The results reveal that uncertainty aversion promotes the separation of semantic representations at both the psychological (experiment 1) and the neural levels (experiment 2). We then tested the direct behavioural consequences of the semantic expansion by examining how uncertainty aversion shapes the trade-off between discriminability and generalization. We find that uncertainty aversion reduces semantic interference (experiment 3) and leads to poorer generalization (experiment 4). In other words, people averse to uncertainty exhibit more distinct semantic representations, which results in them prioritizing the need to discriminate between concepts, at the cost of their ability to generalize.

Results

Uncertainty aversion is associated with a semantic expansion

In experiment 1, 103 participants made semantic relatedness judgements between 16 target words and 42 comparison words (Methods

https://doi.org/10.1038/s41562-023-01561-5

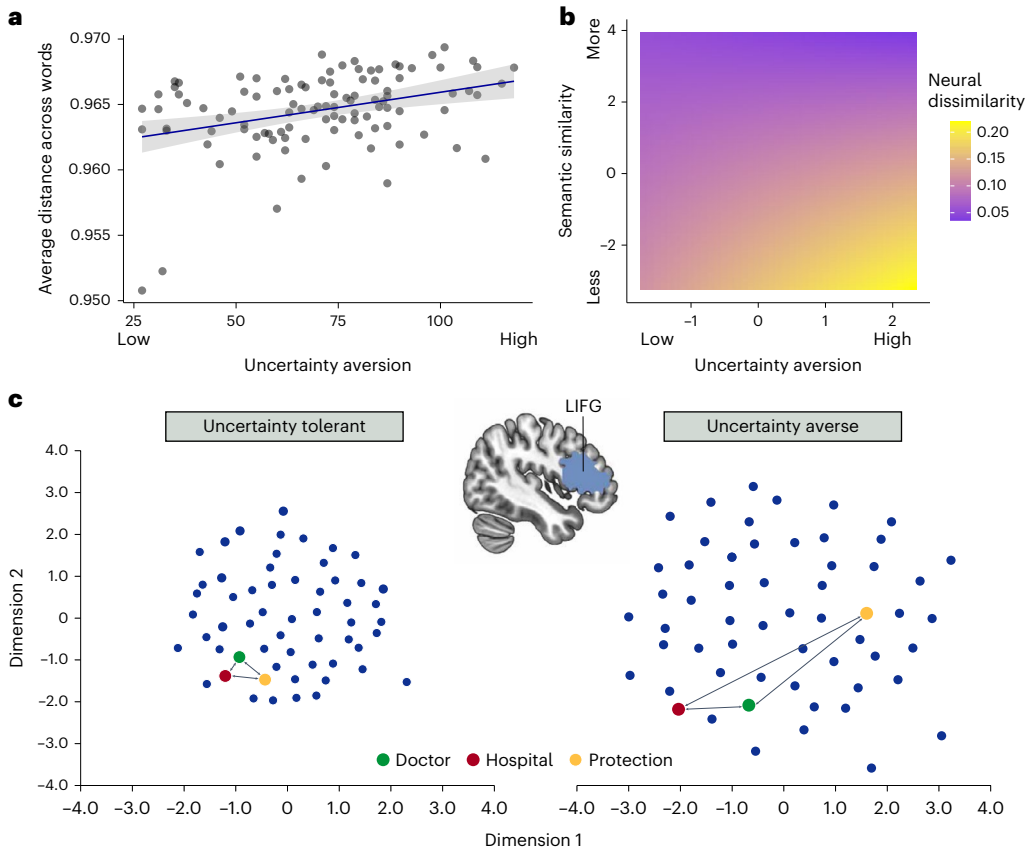

Fig. 2 | The semantic-expansion at the psychological and neural level. **a**, A significant correlation (r) was observed between uncertainty intolerance and distances between semantic representations. The shaded area represents the 95% confidence interval; the result remains significant even after removing the two outliers, $P = 0.01$, two-sided test. **b**, Uncertainty aversion modulates the relationship between semantic similarity and neural similarity. Semantic representations were captured by a reduced neural space for people tolerant of uncertainty, while the neural space was larger for people averse to uncertainty. Cross-validated Mahalanobis distances between pairs of words plotted as a function of beta estimates extracted from the regression model. **c**, This relationship between uncertainty aversion and representational distance between words in the LIFG is further illustrated using two representative participants that differed in their uncertainty attitudes. MDS was applied to their average neural RDMs to transform the neural representational distances for all word pairs into a two-dimensional space. Each word is represented by a point in the plane. Words are more distant from one another for the uncertainty-averse participant.

and Fig. 1). Similarity between the 16 target words was then estimated by correlating the 42-element vectors of relatedness ratings across words. To calculate distances between words using the same number of meaningful dimensions for each participant[25], multidimensional scaling (MDS) was applied. A four-dimensional partition was selected given that it was the lowest number of dimensions that produced a good fit (stress = 0.09)[26]. In this four-dimensional space, the average distance between words was computed for each participant and then correlated with individuals' uncertainty attitudes, which were captured by the well-validated intolerance of uncertainty scale (IUS). The IUS assesses uncertainty aversion by asking people to rate to what extent statements such as 'the ambiguities in life stress me' describe them[9]. Showing a semantic-expansion effect, we found greater aversion to uncertainty was associated with an increase in the distance between concepts in psychological space ($r = 0.35$, $P < 0.001$; Fig. 2a). This relationship was not dependent on the number of dimensions selected when applying MDS since the same result also holds for three, five and six dimensions (Supplementary Table 1). This effect was not driven by an increase in response variability in people with higher IUS scores: there was no significant relationship between response variability and uncertainty aversion ($r = -0.03$, $P = 0.79$; Supplementary Fig. 1), and the semantic-expansion effect remained significant after controlling for response variability in a regression ($\beta = 0.001 \pm 0.0002$ (s.e.), $P < 0.001$). The effect also remains significant after controlling for global variables such as IQ, age and gender ($\beta = 0.001 \pm 0.0003$ (s.e.), $P = 0.002$; Supplementary Table 2).

The expansion of semantic neural representations

To test whether uncertainty aversion is also associated with more distal neural representations, in experiment 2 we analysed data from a functional magnetic resonance imaging (fMRI) experiment in which 44 participants read 60 words and were asked to think about their meaning (Methods). We used representational similarity analysis (RSA) to estimate the distance between words in neural semantic space. Given previous work showing a role for the left inferior frontal gyrus (LIFG), angular gyrus, middle temporal gyrus (MTG), anterior temporal cortex (ATC) and perirhinal cortex in semantic representation[27–32], we hypothesized that the pair-wise distances between the neural representations of words in these brain regions (localized to the left hemisphere[33]) would be sensitive to uncertainty aversion. We further reasoned that the ventromedial prefrontal cortex (VMPFC) as well as the precuneus—regions known to process various forms of uncertainty[6,34–38] and in some cases have also been implicated in semantic processing (that is, the precuneus)[31]—might be involved in indexing the relationship between uncertainty intolerance and how words are represented.

https://doi.org/10.1038/s41562-023-01561-5

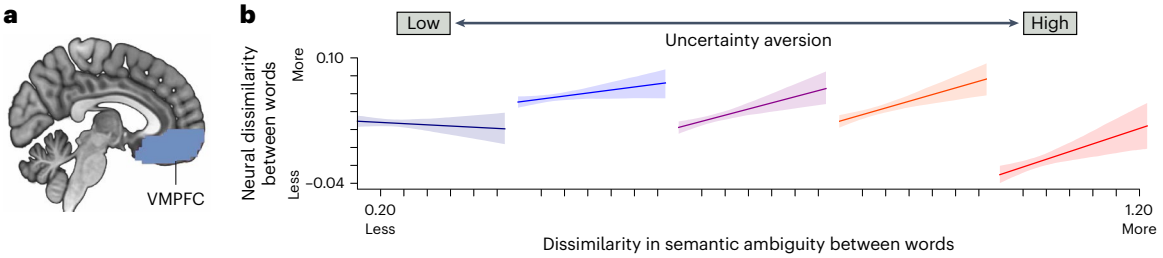

Fig. 3 | People averse to uncertainty encode semantic ambiguity in the VMPFC. a, The VMPFC was a pre-defined ROI. **b**, For visualization purposes, participants were binned into five groups based on discrete IUS intervals. For each participant, a neural RDM was estimated, vectorized and then correlated with the vectorized semantic ambiguity RDM derived from the language corpus. For those averse to uncertainty, activity patterns in the VMPFC are more dissimilar between words that are semantically ambiguous. The shaded area represents the 95% confidence interval.

In any neural semantic representation, the neural pattern distance between a pair of concepts should be proportional to their semantic dissimilarity. We reasoned that, under the semantic-expansion hypothesis, uncertainty aversion would increase the slope of the relationship between semantic dissimilarity and neural representational dissimilarity, as people averse to uncertainty expand their neural semantic space. To test this prediction, for each participant, a neural representational dissimilarity matrix (RDM) was constructed for each region of interest (ROI) by computing the cross-validated Mahalanobis distance between all word pairs[39]. A model RDM was also constructed containing the pair-wise differences of the semantic dissimilarity between words derived from Global Vectors for Word Representation (GloVe; Methods). For each of our ROIs, a linear mixed-effect regression was run with the vectorized lower triangle of the neural RDM as the dependent variable, and the vectorized lower triangle of the semantic similarity RDM, an individual's tolerance to uncertainty, and their interaction, all as predictors. In the LIFG, we observed a significant positive interaction between semantic similarity and aversion to uncertainty: increasingly dissimilar words exhibit increasingly dissimilar activity patterns, especially for individuals who are averse to uncertainty ($\beta = 0.02 \pm 0.006$ (s.e.), $P = 0.005$; Bonferroni-corrected for multiple comparisons across seven ROIs, Fig. 2c). Including age, gender and level of education as covariates did not change this result (Supplementary Results Experiment 2). This effect was not observed in any other ROIs (Supplementary Table 3).

We also tested the possibility that this neural semantic-expansion effect was driven by increased executive control exerted by participants with high IUS scores, which could in theory produce an online reshaping of the semantic representation. To test this, we estimated the mean activity in the fronto-parietal regions of the brain[40], which are known to index executive control and mental effort[40], and included them as covariates in our regression. Again, the significant interaction between semantic similarity and IUS remained significant ($\beta = 0.03 \pm 0.005$, $P < 0.001$).

Uncertainty aversion and semantic ambiguity

The uncertainty associated with multiple semantic representations cued by a given word is higher for words that have multiple meanings and are thus considered semantically ambiguous. For example, 'run' can be used in a diverse array of contexts, whereas 'artichoke' is typically only used in the context of food. A word that can be used in many different contexts has a larger number of referents and greater semantic ambiguity[41]. For those intolerant to uncertainty, encoding this ambiguity would enable semantic representations to be tagged as ones that must be structured more distally. To test this, we used RSA to estimate the pair-wise neural pattern distances between each word. Given that each word is associated with a certain degree of semantic ambiguity—estimated using a corpus that counts the number of contexts a given word appears in[41]—we can interrogate whether our ROIs[6,34,35,37,38,42] track this semantic ambiguity as a function of an individual's aversion to uncertainty. We constructed a neural RDM comprising the pair-wise cross-validated Mahalanobis distances between individual words, and a model RDM containing the pair-wise differences of the semantic ambiguity between words derived from a language corpus (Methods). We then ran a linear mixed-effects regression analysis in which the vectorized neural RDM served as the dependent variable, and the vectorized semantic ambiguity RDM and an individual's tolerance to uncertainty (as well as their interaction) served as predictors for our ROIs.

Given the importance of semantic ambiguity to uncertainty-averse individuals, we expected ambiguity to be encoded more strongly for participants with higher IUS. Indeed, in the VMPFC, we observed a significant positive interaction between semantic ambiguity and aversion to uncertainty: words that are similar in their semantic ambiguity exhibit similar activity patterns, especially for individuals who are averse to uncertainty ($\beta = 0.03 \pm 0.006$ (s.e.), $P < 0.001$; Fig. 3b). A similar effect, albeit of lower magnitude, was also observed in the precuneus and LIFG, as indexed by a significant interaction between intolerance of uncertainty and semantic ambiguity (precuneus: $\beta = 0.02 \pm 0.006$ (s.e.), $P = 0.03$; LIFG: $\beta = 0.02 \pm 0.006$ (s.e.), $P = 0.04$); these effects, however, do not remain significant after controlling for increased cognitive effort presumably associated with processing semantically ambiguous words (see Supplementary Table 4 for the rest of the ROIs).

Uncertainty aversion improves semantic discrimination

The semantic-expansion hypothesis posits that uncertainty-averse participants reduce uncertainty by increasing the separation between concepts in both psychological and neural space. A natural consequence of this hypothesized semantic expansion is that for those intolerant to uncertainty, similar concepts can be more easily discriminated and are less likely to produce semantic interference. To test for this reduction in semantic interference, we leveraged the classic Deese–Roediger–McDermott false memory paradigm in which participants first encode lists of words and are then asked to discriminate between old and new words in a surprise recognition memory test (Fig. 4a). A robustly observed effect within the literature is that lure words semantically similar to the initial words (for example, snow and cold) tend to be falsely endorsed as old[15,16].

In experiment 3, 205 participants first studied 50 words and were then given a surprise recognition memory task with 100 words, 50 of which were the ones previously shown and 50 of which were novel (Methods and Fig. 4a). Participants performed the task reasonably well, correctly identifying the old words 72% of the time (hit rate) while misidentifying the new words 31% of the time (false alarm rate). To test the prediction that uncertainty aversion would modulate the effect of semantic interference, we focused our analysis on the false alarm rate. We first estimated semantic similarity between the novel and previously seen words using an unsupervised learning algorithm trained on a large language corpus (GloVe). The maximum similarity

https://doi.org/10.1038/s41562-023-01561-5

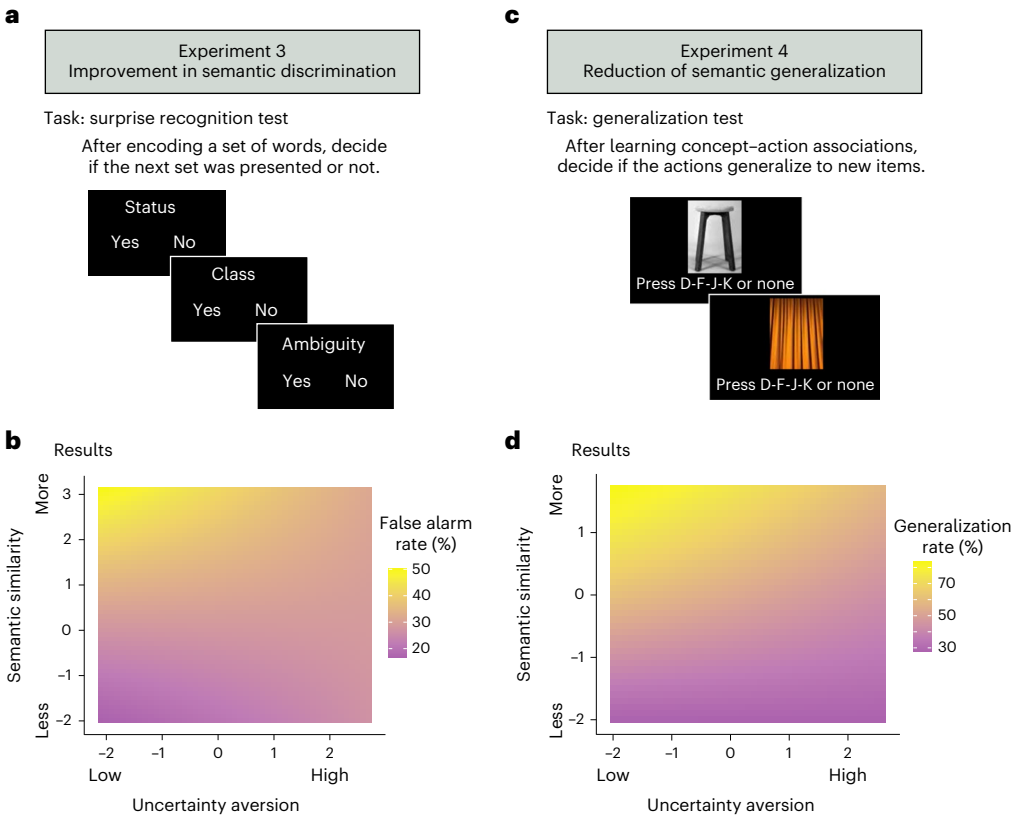

Fig. 4 | Semantic expansion improves discrimination but reduces generalization. a, A surprise recognition test was administered to participants to test for an increase in semantic discrimination for people averse to uncertainty. **b**, Uncertainty aversion modulates the effect of semantic interference during a memory recognition test. While those who can tolerate uncertainty do falsely report recognizing words that are similar to previously presented words, semantic interference is significantly lower for those who are averse to uncertainty. **c**, A generalization task was administered to test for a decrease in semantic generalization for people averse to uncertainty. **d**, Uncertainty aversion modulates the effect of semantic similarity during a generalization task. Semantically closer pictures supported stronger generalization, but this effect was attenuated in those averse to uncertainty. False alarm rates (**b**) and generalization rates (**d**) plotted as a function of beta estimates extracted from respective regression models.

between previously seen words and each new word was calculated and used as a regressor (alongside participants' IUS scores, as well as their interaction) to predict the false alarm rate. Results reveal that as the similarity between novel and previously seen words increases, so too does the probability of falsely identifying a novel word as being already seen ($\beta = -0.22 \pm 0.11$ (s.e.), $P = 0.03$)—an effect that was modulated by one's tolerance to uncertainty (interaction: $\beta = 0.06 \pm 0.02$ (s.e.), $P = 0.01$, Fig. 4b). That is, the greater the aversion to uncertainty is, the better the discrimination between similar concepts. This effect remains significant when adding average hit rate as a regressor in the model ($\beta = 0.06 \pm 0.02$ (s.e.), $P = 0.01$; Supplementary Fig. 2).

Uncertainty aversion reduces semantic generalization

Classic work shows that generalization between stimuli decreases as the psychological distance (that is, discriminability) between the stimuli increases[43]. By this logic, a direct consequence of semantic expansion—which is observed in those averse to uncertainty—should be a reduction in semantic generalization. To test this, in experiment 4, 197 participants first learned an association between 4 concepts (for example, chair and wrench, which were cued by images of objects) and 4 actions (pressing 1 of 4 possible keys). To learn the associations, participants were presented with an image of an object and told to press one of the keys. Feedback about whether the correct key was pressed was immediately provided. Once they learned which concept was associated with which key, they moved to a generalization phase in which they were presented with new pictures of other concepts (for example, stool, screwdriver). This time, they were instructed to press whichever key they saw fit (no feedback was provided during the generalization phase; Methods and Fig. 4c). For each of the four concepts presented during learning, two concepts were selected as stimuli in the generalization phase, one which was semantically close to the initial concept (for example, stool is conceptually similar to chair) and another that was further away (for example, curtain is conceptually distant to chair) but still semantically closer than any of the other target concepts (for example, curtain is even more distant to jellyfish than chair). Pressing the same key associated with the concept during learning was coded as successful generalization (=1), while any other response was coded as lack of generalization (=0). Semantic similarity between the novel concepts during generalization and initial concepts during learning was estimated using GloVe to predict the probability of generalization, alongside participant's aversion to uncertainty and their interaction. As expected, results reveal that the probability of generalization increased as the similarity between novel and target concepts increased ($\beta = -0.83 \pm 0.042$ (s.e.), $P < 0.001$), an effect that was modulated by uncertainty aversion (interaction: $\beta = 0.08 \pm 0.03$ (s.e.), $P = 0.01$), such that semantically similar concepts supported a lower level of generalization as uncertainty aversion increased (Fig. 4d).

Discussion

Uncertainty is present in virtually all aspects of our lives, from perceptual judgements to decision outcomes. A major source of uncertainty comes from deciphering meaning in a conceptually ambiguous world. Not only are most stimuli associated with more than one meaning[44] but also a stimulus can activate multiple similar semantic representations[45]

(for example, slip and lapse), which results in greater uncertainty. Here we directly tested one strategy for reducing this type of uncertainty, which is to expand one's semantic representations.

Behavioural and neural results illustrate the semantic-expansion hypothesis. Aversion to uncertainty scales with more differentiated semantic representations (experiment 1), which is also associated with increasingly separated neural activity patterns in the LIFG (experiment 2)—a region classically involved with matching words with their appropriate semantic representation[46–49]. Two direct behavioural tests of this semantic-expansion hypothesis show that uncertainty-averse individuals exhibit improved semantic discrimination during a surprise memory recognition test (experiment 3) and reduced semantic generalization for similar concepts (experiment 4). As uncertainty becomes more aversive, the structure of semantic representations becomes increasingly separated so that each concept is more easily identified, and as a consequence, the mapping between stimulus and meaning becomes more certain. Together, these findings show that personality needs, such as the need to reduce uncertainty, can shape fundamental aspects of cognition, including how we represent meaning. Indeed, our results suggest that an organizing principle of semantic representations might be the amount of uncertainty people can tolerate.

The fact that we found that the expansion of neural activity patterns was localized to the LIFG—and not a broader neural network associated with uncertainty[50–52]—suggests that regions not traditionally related to uncertainty processing also play a distinct role in helping to reduce semantic uncertainty. Taken a step further, rather than relying on the arbitration of specific neural regions to encode various types of uncertainty[50–52], it is possible that the brain reduces uncertainty by shaping the geometry of neural activity patterns associated with concepts (or stimuli). This account dovetails with current, broader theories of neural activity that posit a relationship between distinct representations and reduced task interference[20,23,24].

Our findings also show that classic uncertainty regions, namely the VMPFC and precuneus, are especially sensitive to semantic uncertainty in those averse to uncertainty—linking the neural encoding of semantic representation to regions involved in processing uncertainty[27]. Considering that the semantic-expansion result was localized to the LIFG, this finding suggests a neural division of labour. In a situation devoid of context in which virtually any meaning associated with each word could be retrieved, VMPFC appears to signal the differences in the number of possible referents that could be retrieved. In contrast, the LIFG plays a role in biasing the structure of semantic representations—through the alteration of the geometry of neural representations—which expands depending on one's aversion to uncertainty, thus avoiding semantic confusion. This accords with previous work linking LIFG activity with the resolution of semantic ambiguity[53].

Even though the specific role of the LIFG in language processing is still debated[54], it has been implicated in semantic control[55,56]. The fact that the observed semantic-expansion effect was localized only to the LIFG may suggest that semantic expansion does not occur in stable semantic representations but is instead produced online based on task demands. We caution, however, against interpreting the LIFG as the sole player involved in processing semantic expansion for two reasons. First, the fact that we did not observe semantic expansion in the ATC may simply be because our study was not optimized to capture the signal in this region. Second, the online account is hard to reconcile with our finding that semantic interference is reduced in those averse to uncertainty, since there is no task demand to discriminate concepts during encoding but only later, during a surprise recognition memory test. Indeed, similar semantic interference effects have previously been associated with medial temporal regions[18] rather than the LIFG. Future work should clarify whether semantic expansion is a property of the stored representation, the result of an online computation or both.

Regardless, the resolution of semantic ambiguity through semantic expansion has behavioural consequences that can be understood through the lens of a computational trade-off between discrimination and generalization[23,43,57]. The world would be impossible to navigate if we had to learn from scratch how to engage with each new stimulus. Instead, we routinely generalize what we learn about from one stimulus to other conceptually similar stimuli[43,57,58]. This sensitivity to the similarity between concepts, however, comes at the cost of increased uncertainty and the possibility of interference. We are more likely to confuse two similar concepts with one another, resulting in errors of discrimination. A natural consequence of semantic expansion is that it shifts this trade-off in favour of discrimination. Increasing the separation of concepts in a psychological and neural semantic space would make concepts more discriminable. We show that people averse to uncertainty are indeed less affected by interference between similar semantic representations and show improved discrimination between concepts. Simultaneously, they are also less able to generalize behaviours between similar concepts. In other words, uncertainty aversion leads to people making their world more compartmentalized, trading off generalization in favour of better discriminability and lower uncertainty.

Important open questions remain. First, does semantic expansion preferentially shape concepts that are closer together or generally influence the whole representation? When the structure of semantic representation was measured (experiments 1 and 2), we found no direct evidence that the expansion effect is greater for similar concepts. However, in experiments 3 and 4, people averse to uncertainty exhibited greater semantic discrimination (experiment 3) and poorer generalization (experiment 4), especially for very similar concepts. It is possible that the different task demands in these latter experiments made our measurements more sensitive for detecting the semantic-expansion effect with concepts that share increasing similarity, and future work can help arbitrate whether this is the case. Second, might our findings be explained by expanding the dimensionality of the representation? Previous theoretical work has argued that the dimensionality of a representation controls a trade-off between discriminability and generalizability[23], and indeed, an increase in dimensionality would result in greater separability. In line with this view, we found that those averse to uncertainty had higher discriminability (and thus lower uncertainty) at the cost of poorer generalizability (experiments 3 and 4). However, we found no direct evidence that uncertainty aversion is associated with higher dimensionality (experiments 1 and 2; Supplementary Figs. 3 and 4). It is possible that, given the low dimensionality of the representational structures that we examined and the poor sensitivity of current dimensionality measures, our analysis was not sufficiently powered to detect higher dimensionality for people averse to uncertainty, which leaves it as an open question for future work.

Research has investigated how diverse experiences alter mental and neural representations, mostly by measuring different levels of expertise (for example, trained musicians versus non-musicians[59]). Rather than showing how different experiences shape neural representations, the semantic-expansion hypothesis proposes that differences can originate even when people share the same experience. Based on the notion that those averse to uncertainty are more likely to adopt strategies that decrease uncertainty and increase identifiability, we have demonstrated that this stable personality trait predicts a more distal structure of semantic representations—a mechanism that could be applied across domains (for example, animals, objects) and cognitive processes (for example, perception).

Methods

All experiments were approved by Brown University's institutional review board. All participants provided informed consent as approved by the institutional review board.

Experiment 1

Participants. A total of 108 people participated in the experiment on Prolific. Five participants were excluded after they failed to pass an

attention check at the end of the experiment, resulting in a final sample of 103 participants (47 women, 51 men, 3 non-binary and 2 transgender; mean age 34 years, s.d. = 12).

Word stimuli. For consistency, we selected a subset of words that were also used in experiment 2, which were originally selected to address an orthogonal research question. Apart from these words, we identified other words that the target words could be compared with. To do this, we searched through lists of abstract words and selected 48 words that we (1) classified as abstract (for example, life, system), (2) were comprehensible and (3) varied in how similar they were to each other. A table with these words can be found in Supplementary Information with their summary characteristics.

Procedure. Participants completed 672 relatedness ratings between 16 target words and 42 comparison words on a scale from 1 (not related) to 9 (very related). The target and comparison words were presented on the screen simultaneously, and participants then judged to what extent they were related. The instructions did not ask participants to focus on a specific semantic domain. Words were selected to be relatively abstract: freedom, family, choice (see full list in Supplementary Information). Following this task, participants' uncertainty attitudes were captured using the 27-item IUS as well as demographic information. The experiment lasted around 40 min, and participants received US$4.50 as compensation.

Experiment 2

Participants. A total of 44 participants, all of whom were right-handed (17 women and 27 men; mean age 32 years, s.d. = 14) participated in the experiment. The data analysed here were collected as part of a larger study exploring the neural mechanisms of intolerance to uncertainty. Participants provided informed consent and received US$40 as monetary compensation.

Procedure. Participants completed a battery of tasks inside and outside the MRI scanner. Before the MRI scanning session began, participants provided basic demographic information and completed a list of questionnaires including the 27-item IUS[9]. The scanning session lasted approximately 1.5 h, in which they completed a word reading task and a video watching task (not analysed in this research). After the scanning session, participants completed other questionnaires and tasks that were also not analysed for this research.

Word reading task. The word reading task was conducted inside the MRI scanner. The task consisted of 6 runs, each made up of 80 trials. In every run, 60 unique words were individually presented on the centre of the screen without any accompanying context. Each trial stimulus was presented for 2.5 s with a fixed inter-stimulus interval of 2.5 s during which only a fixation cross appeared on the screen. Twenty null trials were added in which only a fixation cross was shown to improve the efficiency of the design for estimating stimulus-evoked activity to each word. The trial order was randomized, and the words were presented in black font on a white background screen. To maintain the participant's attention, participants had to indicate if they believed the word presented was political or non-political by pressing one of two response buttons. The total run duration was approximately 6 min, and each participant completed 6 runs.

fMRI acquisition and pre-processing. MR images were collected on a Siemens Prisma Fit 3-Tesla research-dedicated scanner. T2*-weighted functional scans were acquired using a multi-slice sequence capturing three slices at once to ensure whole-brain coverage with short repetition time (T_R = 1,500 ms). Sixty 3 mm transverse slices were acquired, each with 64 × 64 voxels of 3.0 mm isotropic, building up a field of view that covered the entire brain except part of the cerebellum. The field of view was tilted upward by 25° at the front of the brain to minimize tissue gradient-related signal dropout in the orbitofrontal cortex. Contrast settings were optimized for cortical grey matter (T_E(Echo Time) = 30 ms, flip angle = 86°). T1-weighted (T1w) anatomical scans were acquired using a standard MPRAGE sequence (160 sagittal slices with 256 × 256 voxels of 1.0 mm isotropic, T_R = 1,900 ms, T_E = 3.02 ms, flip angle = 9°). Pre-processing was performed using fMRIPrep 1.5.1rc2[60] (RRID:SCR_016216), which is based on Nipype 1.3.0-rc1[60,61] (RRID:SCR_002502).

Anatomical data pre-processing. The T1w image was corrected for intensity non-uniformity with N4BiasFieldCorrection[62], distributed with ANTs 2.2.0[63] (RRID:SCR_004757) and used as T1w reference throughout the workflow. The T1w reference was then skull stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter and grey matter was performed on the brain-extracted T1w image using fast[64] (FSL 5.0.9, RRID:SCR_002823). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both the T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c[65] (RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym).

Functional data pre-processing. The following pre-processing was performed for each of the BOLD runs per participant. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD reference was then co-registered to the T1w reference using FLIRT[66] (FSL 5.0.9) with the boundary-based registration[67] cost function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt[68] (FSL 5.0.9). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207[69] (RRID:SCR_005927). The BOLD time series (including slice-timing correction when applied) were re-sampled onto their original native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These re-sampled BOLD time series referred to pre-processed BOLD. The BOLD time series were re-sampled into standard space, generating a pre-processed BOLD run in MNI152NLin2009cAsym space. Several confounding time series were calculated based on the pre-processed BOLD: frame-wise displacement, DVARS (defined as the temporal derivative of root mean square variance over voxels) and three region-wise global signals. Frame-wise displacement and DVARS are calculated for each functional run, both using their implementations in Nipype[70]. The three global signals are extracted within the CSF, the white matter and the whole-brain masks.

Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor[71]). Principal components are estimated after high-pass filtering the pre-processed BOLD time series (using a discrete cosine filter with 128 s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical grey matter regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and white matter masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the white matter and

CSF masks. For each CompCor decomposition, the first *k* components with the largest singular values are retained, such that the time series of the retained components are sufficient to explain 50% of the variance across the nuisance mask (CSF, white matter, combined or temporal)[72]. The remaining components are dropped from consideration.

The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head-motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each[73]. Frames that exceeded a threshold of 1.0 mm frame-wise displacement were annotated as motion outliers. All re-samplings can be performed with a single interpolation step by composing all the pertinent transformations (that is, head-motion transform matrices and co-registrations to anatomical and output spaces). The data were not susceptibility-distortion corrected in the absence of field maps. The gridded (volumetric) re-samplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels[74]. Non-gridded (surface) re-samplings were performed using mri_vol2surf (FreeSurfer).

fMRI data cleaning. For the word reading task, two exclusion criteria were used. First, we excluded runs in which there were excessive motion artefacts, defined as a frame-wise displacement greater than 1 mm in more than 10% of the repetition times. Second, runs in which participants did not press the response button because of experimenter error or a lack of attention were also excluded. This resulted in a total exclusion of 8 runs from 6 participants (on average 1.3 runs per participant and not more than 2 runs for each participant, which leaves at least 4 presentations of each word for each participant) and the exclusion of one participant all together, leading to a final inclusion of 43 participants for the statistical fMRI analyses.

Statistical fMRI analysis. ROI. Based on previous research on semantic processing[27–32], we defined our ROIs as the LIFG, angular gyrus, MTG, ATC and perirhinal cortex. The LIFG was defined using the triangular and opercular part of the LIFG, as stipulated by the Automated Anatomical Labeling (AAL) atlas[75]. The AAL atlas was also used to demarcate the angular gyrus, MTG and ATC. The perirhinal cortex was defined as Brodmann areas 35 and 36. We also had a priori ROIs related to uncertainty[6,34–38], which included the VMPFC and precuneus—which were also demarcated using the AAL.

RSA. For every participant, a general linear model was constructed using SPM12 in MatlabR2017b, in which every word that was presented in the word reading task was modelled as a separate regressor yoked the duration of stimulus display. Motion was regressed using six motion directions, their first derivatives, their squares and the first derivatives of the squares. In addition, the CompCor physiological regressors described above were included as covariates for de-noising. Fitting this model to the data yielded a beta map per run, per word. The beta values for each voxel were extracted for each ROI. The RSA Toolbox[76] (http://github.com/rsagroup/rsatoolbox) was then used to compute, for each participant and ROI, a neural RDM using the cross-validated Mahalanobis distance between all word pairs. This measure incorporates multivariate noise normalization and cross-validation, which leads to higher reliability of the neural distance estimates[39].

Testing for semantic expansion. To investigate whether intolerance of uncertainty expands the neural representation of semantics, we adopted a linear mixed-effects model in which the neural representational dissimilarity of each word pair was regressed onto uncertainty attitudes and semantic dissimilarity as captured by GloVe. A neural RDM for each ROI was computed using the approach described above. A model RDM containing the pair-wise dissimilarity in semantic dissimilarity was created by computing the difference between semantic similarity scores obtained from GloVe. Both neural and model RDMs were vectorized and *z*-scored. The main effects of IUS and semantic dissimilarity and the interaction between the two were tested with the participant included as a random intercept[77].

Semantic ambiguity analysis. To investigate whether intolerance of uncertainty modulates how semantic ambiguity is processed, we adopted a linear mixed-effects model in which the neural representational dissimilarity of each word pair was regressed by uncertainty attitudes and semantic ambiguity. A neural RDM for each ROI was computed using the approach described above. A model RDM containing the pair-wise dissimilarity in semantic ambiguity was created by computing the difference between semantic ambiguity scores obtained from the English Lexicon Project Web Site[78] (https://elexicon.wustl.edu/). Both neural and model RDMs were vectorized and z-scored. The main effects of IUS and semantic ambiguity and the interaction between the two were tested with the participant included as a random intercept[77].

Experiment 3

Participants. A total of 213 participants were run on Prolific; 7 participants were removed as they failed to pass an attention check at the end of the experiment, and 1 participant because they did not complete the whole experiment, resulting in a total of 205 participants (103 women, 96 men, 4 non-binary and 2 gender fluid; mean age 29 years, s.d. = 6).

Word stimuli. Concepts were selected to be relatively abstract and include a mix of relatively similar (for example, status–class) and dissimilar (for example, ambiguity–peculiarity) new and old words.

Procedure. Participants were first presented with a word and a fractal and were told that these two stimuli were associated. After this, participants were asked to judge the likelihood that 10 other words were also associated with the same fractal on a scale from 1 (not likely) to 7 (very likely). Participants completed 5 blocks of this task. At the beginning of each block, a new fractal was associated with a new word, and participants had to judge the association between the fractal and 10 new, comparison words. In total, they completed 50 trials of this task for a total presentation of 50 comparison words. A surprise recognition test was then administered. In total, participants were presented with 100 words (50 old and 50 foils; for a subset of participants, a coding error made the presented foils to be only 48) and had to indicate whether they had seen this word before (Yes or No). Finally, participants were asked to judge the similarity between the 50 word pairs that they saw in the associative task on a scale from 1 (not similar) to 7 (very similar). Again, uncertainty attitudes were collected using the IUS scale together with demographic information. The experiment lasted around 15 min, and participants received US$2 as compensation.

Analysis procedure. A mixed-effect binomial regression was performed on the probability of falsely recognizing a lure (1 = yes, 0 = no) predicted by the maximum similarity of the old words for each lure, uncertainty attitudes and the interaction between the two. Subject and word item were both added as independent random intercepts.

Experiment 4

Participants. A total of 221 participants were tested on Prolific; 13 participants were removed as they failed to pass an attention check at the end of the experiment, 9 participants were removed because their average accuracy in the learning phase was lower than 70% and 2 participants were removed because their average reaction time in the generalization task was lower than 500 ms, resulting in a total of 197 participants (94 women, 97 men, 6 non-binary; mean age 29 years, s.d. = 6).

Procedure. The experiment consisted of two phases, a learning phase and a generalization phase. In the learning phase, participants learned the association between four pictures cueing different concepts (for example, jellyfish, chair, canoe) and four keys (for example, 'd', 'f', 'j', 'k'). On each trial, a picture was presented and participants pressed one of the four keys. Immediately after, participants received feedback regarding whether the selected key was the one associated with the concept. Participants responded to 20 trials for each associated pair in a pseudo-randomized order. After learning, participants completed a generalization phase, in which new pictures of other concepts were presented and participants had to decide whether to select one of the old keys or the space bar that represented a 'no response'. Two concepts were selected for each of the four target concepts, one that was very semantically close to the target (for example, chair–stool) and another one that was semantically further (for example, chair–curtain), for a total of eight pictures depicting eight different concepts. Each picture was presented twice in a pseudo-randomized order. In total, participants completed two independent blocks of learning and generalization for a total of eight pair associates (see Supplementary Information for the list of concepts presented). The presentation order was counterbalanced across participants. At the end of the experiment, uncertainty attitudes were collected using the IUS scale together with demographic information. The experiment lasted around 15 min, and participants received US$2 as compensation.

Analysis procedure. A mixed-effect binomial regression was performed on the probability of generalization (1 = yes, 0 = no) predicted by the similarity of the new concept with the target concept, uncertainty attitudes and the interaction between the two. Subject and category were both added as independent random intercepts.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

De-identified data for all experiments are publicly available at https://osf.io/gmqh4. The data used for analyses to estimate semantic similarity and ambiguity are available at https://nlp.stanford.edu/projects/glove/ and https://elexicon.wustl.edu/.

Code availability

Codes for the analyses of the paper are publicly available at https://osf.io/gmqh4.

References

1. Kahneman, D. & Tversky, A. Variants of uncertainty. *Cognition* **11**, 143–157 (1982).
2. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13**, 572–586 (2012).
3. Vives, M. L. & FeldmanHall, O. Tolerance to ambiguous uncertainty predicts prosocial behavior. *Nat. Commun.* **9**, 2156 (2018).
4. Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263 (1979).
5. FeldmanHall, O. & Shenhav, A. Resolving uncertainty in a social world. *Nat. Hum. Behav.* **3**, 426–435 (2019).
6. Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T. & Platt, M. L. Neural signatures of economic preferences for risk and ambiguity. *Neuron* **49**, 765–775 (2006).
7. Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J. & Ladouceur, R. Why do people worry?. *Pers. Individ. Differ.* https://doi.org/10.1016/0191-8869(94)90048-5 (1994).
8. Hirsh, J. B., Mar, R. A. & Peterson, J. B. Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychol. Rev.* **119**, 304–320 (2012).
9. Buhr, K. & Dugas, M. J. The intolerance of uncertainty scale: psychometric properties of the English version. *Behav. Res. Ther.* **40**, 931–945 (2002).
10. Ladouceur, R., Talbot, F. & Dugas, M. J. Behavioral expressions of intolerance of uncertainty in worry: experimental findings. *Behav. Modif.* **21**, 355–371 (1997).
11. Dugas, M. J. et al. Intolerance of uncertainty and information processing: evidence of biased recall and interpretations. *Cogn. Ther. Res.* **29**, 57–70 (2005).
12. Luhmann, C. C., Ishida, K. & Hajcak, G. Intolerance of uncertainty and decisions about delayed, probabilistic rewards. *Behav. Ther.* **42**, 378–386 (2011).
13. Tanovic, E., Gee, D. G. & Joormann, J. Intolerance of uncertainty: neural and psychophysiological correlates of the perception of uncertainty as threatening. *Clin. Psychol. Rev.* **60**, 87–99 (2018).
14. Gilboa, I. & Schmeidler, D. Maxmin expected utility with non-unique prior. *J. Math. Econ.* **18**, 141–153 (1989).
15. Van Baar, J. M., Halpern, D. J. & FeldmanHall, O. Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proc. Natl Acad. Sci. USA* **118**, e2022491118 (2021).
16. Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. Topics in semantic representation. *Psychol. Rev.* **114**, 211–244 (2007).
17. Henderson, J. M. & Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **1**, 743–747 (2017).
18. Chadwick, M. J. et al. Semantic representations in the temporal pole predict false memories. *Proc. Natl Acad. Sci. USA* **113**, 10180–10185 (2016).
19. McEvoy, C. L., Nelson, D. L. & Komatsu, T. What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**, 1177–1194 (1999).
20. Badre, D., Bhandari, A., Keglovits, H. & Kikumoto, A. The dimensionality of neural representations for control. *Curr. Opin. Behav. Sci.* **38**, 20–28 (2021).
21. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
22. Diedrichsen, J., Wiestler, T. & Ejaz, N. A multivariate method to determine the dimensionality of neural representation from population activity. *Neuroimage* **76**, 225–235 (2013).
23. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
24. Franconeri, S. L., Alvarez, G. A. & Cavanagh, P. Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn. Sci.* **17**, 134–141 (2013).
25. Landauer, T. K. & Dumais, S. T. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
26. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).
27. Friederici, A. D. The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392 (2011).
28. Poldrack, R. A. et al. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* **10**, 15–35 (1999).
29. Frankland, S. M. & Greene, J. D. Concepts and compositionality: in search of the brain's language of thought. *Annu. Rev. Psychol.* **71**, 273–303 (2020).

30. Fedorenko, E., Blank, I. A., Siegelman, M. & Mineroff, Z. Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition* **203**, 104348 (2020).
31. Fairhall, S. L. & Caramazza, A. Brain regions that represent amodal conceptual knowledge. *J. Neurosci.* **33**, 10552–10558 (2013).
32. Hoffman, P., Pobric, G., Drakesmith, M. & Lambon Ralph, M. A. Posterior middle temporal gyrus is involved in verbal and non-verbal semantic cognition: evidence from rTMS. *Aphasiology* **26**, 1119–1130 (2012).
33. Knecht, S. et al. Language lateralization in healthy right-handers. *Brain* **123**, 74–81 (2000).
34. Bechara, A., Damasio, H. & Damasio, A. R. Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* **10**, 295–307 (2000).
35. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* **29**, 537–545 (2001).
36. FeldmanHall, O. et al. Stimulus generalization as a mechanism for learning to trust. *Proc. Natl Acad. Sci. USA* **115**, E1690–E1697 (2018).
37. Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. & Camerer, C. F. Neural systems responding to degrees of uncertainty in human decision-making. *Science* **310**, 1680–1683 (2005).
38. Levy, I., Snell, J., Nelson, A. J., Rustichini, A. & Glimcher, P. W. Neural representation of subjective value under risk and ambiguity. *J. Neurophysiol.* **103**, 1036–1047 (2010).
39. Walther, A. et al. Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).
40. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl Acad. Sci. USA* **110**, 16616–16621 (2013).
41. Hoffman, P., Lambon Ralph, M. A. & Rogers, T. T. Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behav. Res. Methods* **45**, 718–730 (2013).
42. FeldmanHall, O. et al. Stimulus generalization as a mechanism for learning to trust. *Proc. Natl Acad. Sci. USA* **115**, 1690 (2018).
43. Shepard, R. N. Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
44. Rodd, J. M., Gaskell, M. G. & Marslen-Wilson, W. D. Modelling the effects of semantic ambiguity in word recognition. *Cogn.Sci.* https://doi.org/10.1016/j.cogsci.2003.08.002 (2004).
45. Roelofs, A. A spreading-activation theory of lemma retrieval in speaking. *Cognition* **42**, 107–142 (1992).
46. Schnur, T. T. et al. Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proc. Natl Acad. Sci. USA* **106**, 322–327 (2009).
47. Vuong, L. C. & Martin, R. C. LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain Lang.* **116**, 22–32 (2011).
48. Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T. & Dronkers, N. F. Specifying the role of the left prefrontal cortex in word selection. *Brain Lang.* **149**, 135–147 (2015).
49. Pino, D., Mädebach, A., Jescheniak, J. D., Regenbrecht, F. & Obrig, H. BONEs not CATs attract DOGs: semantic context effects for picture naming in the lesioned language network. *Neuroimage* **246**, 118767 (2022).
50. Volz, K. G., Schubotz, R. I. & Von Cramon, D. Y. Predicting events of varying probability: uncertainty investigated by fMRI. *Neuroimage* **19**, 271–280 (2003).
51. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
52. Philiastides, M. G., Ratcliff, R. & Sajda, P. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* **26**, 8965–8975 (2006).
53. Grindrod, C. M., Bilenko, N. Y., Myers, E. B. & Blumstein, S. E. The role of the left inferior frontal gyrus in implicit semantic competition and selection: an event-related fMRI study. *Brain Res.* **1229**, 167–178 (2008).
54. Fedorenko, E. & Blank, I. A. Broca's area is not a natural kind. *Trends Cogn. Sci.* **24**, 270–284 (2020).
55. Badre, D. & Wagner, A. D. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* **45**, 2883–2901 (2007).
56. Hoffman, P., McClelland, J. L. & Lambon Ralph, M. A. Concepts, control, and context: a connectionist account of normal and disordered semantic cognition. *Psychol. Rev.* **125**, 293–328 (2018).
57. Sims, C. R. Efficient coding explains the universal law of generalization in human perception. *Science* **360**, 652–656 (2018).
58. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140 (1962).
59. Bangert, M. et al. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage* **30**, 917–926 (2006).
60. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
61. Gorgolewski, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
62. Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
63. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
64. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
65. Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102 (2009).
66. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
67. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
68. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
69. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10**, 171–178 (1997).
70. Power, J. D. et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
71. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
72. Muschelli, J. et al. Reduction of motion-related artifacts in resting state fMRI using aCompCor. *Neuroimage* **96**, 22–35 (2014).
73. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).

74. Lanczos, C. Evaluation of noisy data. *J. Soc. Ind. Appl. Math. B* **1**, 76–85 (1964).
75. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).
76. Nili, H. et al. A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).
77. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
78. Balota, D. A. et al. The English lexicon project. *Behav. Res. Methods* **39**, 445–459 (2007).

Acknowledgements

This work was funded by a Carney Innovation Grant from the Robert J. and Nancy D. Carney Institute for Brain Science and NIH Centers of Biomedical Research Excellence Grant, no. P20GM103645 awarded to O.F.H. A.B. was supported by grant number R01MH125497 from the National Institute of Mental Health and grant number R21NS108380 from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

M.-L.V., D.d.B., O.F.H. & A.B. contributed to designing the research and writing the paper. J.M.v.B performed data collection. M.-L.V. and D.d.B. performed the data analyses.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01561-5.

Correspondence and requests for materials should be addressed to Marc-Lluís Vives, Oriel FeldmanHall or Apoorva Bhandari.

Peer review information *Nature Human Behaviour* thanks Rebecca Jackson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Marc-Lluís Vives
Oriel FeldmanHall
Apoorva Bhandari

Last updated by author(s): Feb 8, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Experiment 1: java script (jsPsych library), cognition.run, and qualtrics. Experiment 2: Psychtoolbox Matlab R2017b, custom Python code. Experiment 3: java script (jsPsych library, 6.3.0), cognition.run, and qualtrics. Experiment 4: OpenSesame (3.3.12) and qualtrics.
Data analysis	Experiment 1: R (4.1.0). Experiment 2: Python (3.7.4), Matlab R2017b with SPM12 and RSA toolbox (https://github.com/rsagroup/rsatoolbox_matlab), R (3.4.3). Experiment 3: R (4.1.0). Experiment 4: R (4.1.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data analysed in this paper are available at https://osf.io/gmqh4. The data used for analyses to estimate semantic similarity and ambiguity are available at https://nlp.stanford.edu/projects/glove/ and https://elexicon.wustl.edu/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	All studies are quantitative. Experiment 1: participants were asked to judge the semantic relatedness of a list of words to estimate their semantic representation. Experiment 2: participants were asked to read words while undergoing functional neuroimaging. Experiment 3: participants were presented a list of words and were surprised with a recognition memory test. Experiment 4: participants learned associations between concepts and keys and were then asked to generalize for other set of concepts.
Research sample	Experiments 1: Adult population of United States collected through Prolific. 103 participants (47 women, 51 men, 3 non-binary, and 2 transgender; mean age 34, SD = 12). The sample is not representative. The sample size was decided to detect an effect size around r=0.2. Experiment 2: Adult population from Rhode Island area. 44 participants, (17 women and 27 men and; mean age 32, SD = 14). The sample is not representative. The sample size was decided to have a sufficient sample to detect differences across ideological groups. Experiment 3: Adult population of United States collected through Prolific. 205 participants (103 women, 96 men, 4 non-binary, and 2 gender fluid; mean age 29, SD = 6). The sample size is not representative. The sample size was decided to detect an effect similar to Experiment 1, but this time the predicted effect was an interaction, so sample size was doubled. Experiment 4: Adult population of United States collected through Prolific. 197 participants (94 women, 97 men, 6 non-binary; mean age 29, SD = 6). The sample size was decided to be the same as in Experiment 3.
Sampling strategy	All experiments used random sampling. Experiment 1: We were expecting a medium to small correlation (0.4-0.2), which requires a sample size of 100 for acceptable power (80%). Experiment 2: Since we were interested in individual differences in neural activity patterns, a relatively large sample was collected (N=44) to test for an orthogonal research hypothesis related to uncertainty attitudes and political polarization. Thus, this sample size was chosen as a result of a trade-off between obtaining reliable estimates of neural activity patterns for each participant while obtaining a sufficiently large sample, given budget limitations. Experiment 3: Based on the effect size of Experiment 1, we expected again a medium to small effect size. Since this time the main prediction was an interaction, and the effect size that we expected to modulate was unknown, we decided to be conversative and double the amount of participants tested for an N=200. Experiment 4: Since we were expecting a similar effect as in Experiment 3, we went for the same sample size.
Data collection	Experiment 1: Computer. It was conducted online, no researcher present during data collection. Experiment was run online until the study sample was completed. The researcher who designed the study was not blinded to the experimental condition nor the study hypothesis. Experiment 2: Computer and magnetic resonance imaging scanner. Research assitants were also present during data collection. Everyone was blind to the hypothesis. Experiment 3: Computer. It was conducted online, no researcher present during data collection. The researcher who designed the study was blinded to the experimental condition and the study hypothesis. Experiment 4: Computer. It was conducted online, no researcher present during data collection. The researcher who designed the study was not blinded to the experimental condition nor the study hypothesis.
Timing	Experiment 1: november 23 2021-november 24 2021 Experiment 2: april 22 2019-october 30 2019 Experiment 3: march 21 2022-march 30 2022 Experiment 4: october 4 2022-october 12 2022
Data exclusions	All exclusion criteria were pre-established.

Data exclusions	Experiment 1: 5 participants were excluded after they failed to pass an attention check at the end of the experiment. Experiment 2: 1 participant for excessive motion artifacts. Experiment 3: 7 participants were removed since they failed to pass an attention check at the end of the experiment, and 1 participant because they did not complete the whole experiment. Experiment 4: 13 participants were removed since they failed to pass an attention check at the end of the experiment, 9 participants were removed because their average accuracy in the learning phase was lower than 70%, and 2 participants were removed because their average reaction time in the generalization task was lower than 500 milliseconds.
Non-participation	No participants dropped out or declined participation.
Randomization	Participants were not allocated into different experimental groups (all conditions within-subjects).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Experiment 1: Ad made available on Prolifics. No self-selection bias anticipated. Experiment 2: Participants were selected to meet certain criteria in terms of politic affiliation for a a larger study exploring the neural mechanisms of intolerance to uncertainty. No direct interference is expected for the main research question explored here. Experiment 3: Ad on Prolifics. No self-selection bias anticipated. Experiment 4: Ad on Prolifics. No self-selection bias anticipated.
Ethics oversight	The study protocol was approved by Brown University's Institutional Review Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Event-related
Design specifications	The task consisted of 6 runs, each made up of 80 trials. In every run, 60 unique words were individually presented on the center of the screen without any accompanying context. Each trial stimulus was presented for 2.5s with a fixed inter stimulus interval of 2.5s during which only a fixation cross appeared on the screen. 20 null-trials were added in which only a fixation cross was shown to improve the efficiency of the design for estimating stimulus evoked activity to each word. Total run duration was approximately 6 minutes and each participant completed 6 runs.
Behavioral performance measures	To maintain participant's attention, participants had to indicate if they believed the word presented was political or non-political by pressing one of two response buttons.

Acquisition

Imaging type(s)	Functional
Field strength	3T
Sequence & imaging parameters	T2*-weighted functional scans were acquired using a multi-slice sequence capturing three slices at once to ensure whole-brain coverage with short repetition time (TR = 1500 ms). 60 3-mm transverse slices were acquired, each with

64x64 voxels of 3.0 mm isotropic, building up a field of view (FOV) that covered the entire brain except part of the cerebellum. The FOV was tilted upward by 25 degrees at the front of the brain to minimize tissue gradient-related signal dropout in the orbitofrontal cortex. Contrast settings were optimized for cortical grey matter (TE = 30 ms, flip angle = 86°). T1-weighted anatomical scans were acquired using a standard MPRAGE sequence (160 sagittal slices with 256x256 voxels of 1.0 mm isotropic, TR = 1900 ms, TE = 3.02 ms, flip angle = 9°).

Area of acquisition

Whole brain scan was used except cerebellum

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software

Preprocessing was performed using fMRIPrep 1.5.1rc254 (RRID:SCR_016216), which is based on Nipype 1.3.0-rc1 (RRID:SCR_002502). The BOLD reference was co-registered to the T1w reference using flirt60 (FSL 5.0.9) with the boundary-based registration61 cost-function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt62 (FSL 5.0.9). BOLD runs were slice-time corrected using 3dTshift from AFNI 2016020763 (RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series are referred here as preprocessed BOLD. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype64. The three global signals are extracted within the CSF, the WM, and the whole-brain masks.

Normalization

Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template.

Normalization template

The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c

Noise and artifact removal

Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor65). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration.

The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each66. Frames that exceeded a threshold of 1.0 mm FD were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices and co-registrations to anatomical and output spaces). The data was not susceptibility distortion corrected in the absence of fieldmaps. The Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels67. Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Volume censoring

A reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep.

Statistical modeling & inference

Model type and settings

Representational Similarity Analysis (RSA). For every participant, a General Linear Model (GLM) was constructed using SPM12 in MatlabR2017b, in which every word that was presented in the word reading task was modelled as a separate regressor yoked the duration of stimulus display, and motion was regressed using 6 motion directions, their first derivatives, their squares, and the first derivatives of the squares. This yielded a beta map per run, per word, thus mapping its neural representation. The beta values in the ROI voxels were extracted for each ROI. The RSA Toolbox69 (http://github.com/rsagroup/rsatoolbox) was then used to compute, for each participant and ROI, a neural representation dissimilarity matrix (RDM) using the cross-validated Mahalanobis distance between all word pairs. This measure incorporates multivariate noise normalization and cross-validation, which leads to higher reliability of the neural distance estimates.

Effect(s) tested

Two main effects were tested:
1. Does a personality trait (intolerance to uncertainty) predict overall dissimilarity in the neural activity patterns elicited while reading words? To test this, we applied multidimensional scaling (MDS) to a representational dissimilarity matrix estimated using RSA (see above). Overall dissimilarity among neural activity patterns for each word was computed by averaging neural distances on the 2-dimensional space elicited by MDS. This measure was then correlated with participants' aversion to uncertainty, which was estimated using a classic scale from past literature (see below).
2. Does a personality trait (aversion to uncertainty) modulate the neural encoding of semantic ambiguity? To test this, we ran a linear mixed-effects model in which the neural representational dissimilarity of each word pair was regressed onto a set of

regressors, which included participant's aversion to uncertainty and semantic ambiguity.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s) All Regions of Interested were demarcated using the Automated Anatomical Labeling atlas.

Statistic type for inference (See Eklund et al. 2016)

We did not carry out any whole-brain univariate or multivariate analyses that required voxel-wise or cluster-wise statistical inference. All our fMRI analyses focussed on a small number of predefined regions of interest which were analyzed using a) representational similarity analysis b) mixed-effects regression.

Correction

We applied bonferroni-corrections for multiple comparisons across seven ROIs.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

Main independent variables:
- Intolerance to uncertainty was measured using the well-validated intolerance of uncertainty scale (IUS). The IUS assesses uncertainty aversion by asking questions like "the ambiguities in life stress me". This variable was then used to predict neural expansion (main effect 1, see above), and the modulation of semantic ambiguity (main effect 2, see above).
- Semantic ambiguity was measured by computing the difference between semantic ambiguity scores of each word obtained from the English Lexicon Project Web Site (https://elexicon.wustl.edu/). This model RDM was then vectorized, z-scored, and included in the mixed-effect regression model (main effect 2, see above).