Supplementary Materials for Zombie Intuitions

Appendix A – Corpus studies (Studies 1A-1C)

Study 1A considered collocates in the *Corpus of Contemporary American English* (COCA), starting with the top 20 noun lemmas that occur most frequently directly after 'zombie'.[1] These terms are highly suggestive of Hollywood zombies, with highly used phrases including 'zombie apocalypse', 'zombie movie', 'zombie attack', and 'zombie plague'. This includes nouns like 'mode' and 'walk' that are not immediately suggestive of zombies, but that indicate the Hollywood sense in context: many popular video games have a 'zombie mode' where you fight the walking dead and a 'zombie walk' is a public gathering where people dress up as Hollywood zombies and walk through public spaces. We next looked at the most frequent lemmas to occur within four words of 'zombie'. As seen in Table A-1, these include nouns like 'apocalypse' and 'movie', adjective like 'mindless' and 'flesh-eating', and verbs like 'bite' and 'attack'. These findings suggest the Hollywood sense of 'zombie' is very prominent.

Table A-1

Top 20 nouns, adjectives, and verbs, within four words of 'zombie' in COCA.

Nouns	Count	Adjectives	Count	Verbs	Count
apocalypse	373	Mindless	47	Bite	64
Movie	212	flesh-eating	35	Attack	59
Film	172	Scary	27	Chase	21
Vampire	107	Outer	22	Flick	18
Survival	93	post-apocalyptic	15	Crawl	14
Horde	88	Nazi	14	Overrun	14
Brain	75	Pandemic	13	Shamble	14
Outbreak	61	Millennial	13	Slay	13
Human	52	Creepy	12	Infect	12
Guide	48	Oral	11	Stab	9
Killer	47	Philosophical	11	Roam	9
Mode	44	Resident	10	Outrun	8
Horror	42	Infected	7	Waltz	7
Dragon	42	real-life	7	Unleash	6
Plague	42	full-on	6	Swarm	6
Virus	41	Brainless	6	Lurch	6
Jesus	38	Brainwashed	6	Geeks	6
Novel	37	mixed-up	5	Chomp	6
Invasion	35	Supernatural	5	Undead	6
Genre	34	Apocalyptic	5	Moan	5

Study 1B examined the dominant sense of 'zombie' as revealed by a distributional semantic model (DSM) built from COCA, excluding academic sources. We employed a context-

[1] Top 20 noun lemmas are: 'apocalypse' (count: 364), 'movie' (333), 'film (121), 'survival' (71), 'attack' (57), 'game' (54), 'outbreak' (53), 'horde' (51), 'story' (55), 'plague' (33), 'war' (32), 'mode' (29), 'killer' (29), 'invasion' (28), 'book' (28), 'walk' (27), 'virus' (25), 'bank' (23), 'novel' (21), 'army' (21).

predicting model generated using the word2vec algorithm with standard parameters and a five-word context window. This model has been previously shown to score extremely well on the MEN benchmark with a Spearman's rho of 0.80 (see Sytsma et al., 2019, for details on DSM and the specific models used). The basic idea behind DSMs follows Firth's dictum that 'you shall know a word by the company it keeps' (Firth, 1957, p.11). DSMs look at the company that each word in a corpus keeps—that is, the terms that occur in proximity to it—and assumes that terms that keep similar company have similar meanings. This information is used to represent terms as vectors in a semantic space. The terms can then be compared to give a measure of similarity of meaning, typically by taking the cosine of the vectors. For context-predicting models, this information is used to train an artificial neural network to set vector weights. These models outperform more traditional DSMs (Baroni et al., 2014).

We looked at the nearest neighbours of 'zombie' in the DSM (the terms with the largest cosine values with it and, hence, the terms that are closest in meaning to it according to the model). The nearest neighbours are suggestive of Hollywood monsters, as seen in Table A-2, indicating that the Hollywood sense of 'zombie' is the dominant sense.

Table A-2

Nearest neighbours to 'zombie' in the non-academic COCA DSM created by Sytsma et al. (2019), showing comparison term and cosine value.

Term	Cosine	Term	Cosine
'undead'	0.67	'crazed'	0.55
'vampire'	0.64	'slayer'	0.55
'ghoul'	0.62	'bloodsucking'	0.55
'Dracula'	0.59	'deranged'	0.53
'buffy'	0.58	'ghost'	0.53

Study 1C assessed the linguistic salience of different senses of 'zombie' more directly. Salience is a function of exposure frequency and prototypicality. We used occurrence frequencies in COCA as proxy for exposure frequency and production frequencies as measure of prototypicality (Chang, 1986). To assess occurrence frequencies, we used a random sample of 500 sentences with the word 'zombie' drawn from COCA. To assess prototypicality, we used a production task and recruited 50 participants using the same method and restrictions as in our main studies.[2] Participants were asked 'Write down 10 sentences that use the word "zombie" in the spaces below. Please try to give varied responses!' We thus obtained 500 produced 'zombie'-sentences. Two independent coders assessed the sentences from the two corpora. The coders were MA students, native speakers of English, ignorant of our research questions. They classified occurrences of the noun 'zombie' as uses of one of 10 senses; where they felt unable to do so, they either indicated insufficient context, that 'zombie' was mentioned rather than used, or that it was ambiguous between Hollywood and voodoo sense (1 or 2 below). The 10 senses were compiled by starting with senses attested by *WordNet*, then adding any further senses and information from *Oxford Dictionaries Online, Oxford English Dictionary*, and *Webster's New World College Dictionary* (see Table A-3).[3]

[2] These participants were 76.0% women, mean age 44.2 years (16-78 years).

[3] https://www.lexico.com/, https://www.oed.com/, https://www.merriam-webster.com/, last accessed 25/7/2019.

Table A-3

Coder evaluations for classifiable sentences in Study 1C.

Sense	COCA			Produced		
	C1	C2	AVG	C1	C2	AVG
1. ‘a corpse reanimated by a virus or other pathogen, typically capable of movement but not of rational thought, and feeding on human flesh, *according to popular fiction and horror movies’*	82.4%	79.8%	81.1%	82.3%	82.3%	82.3%
2. ‘a dead body that has been brought back to life, by a supernatural force, *according to voodoo religion*’	0.3%	0.3%	0.3%	0%	0%	0%
3. ‘a spirit or supernatural force that reanimates a dead body, *according to voodoo religion or stories*’	0%	0%	0%	0%	0%	0%
4. ‘a snake god worshipped by voodoo cults of African origin’	0%	0%	0%	0%	0%	0%
5. ‘a person who is listless or lethargic and acts or responds in a mechanical or apathetic way or is dull, slow-witted’.	12.9%	16.1%	14.5%	15.3%	16.1%	15.7%
6. ‘a weird, eccentric, or unattractive person’, or general term of disparagement	0.9%	0%	0.4%	1.6%	0.4%	1.0%
7. ‘a drink made with several kinds of rum, liqueur, and fruit juice’	2.3%	2.6%	2.5%	0.8%	0.8%	0.8%
8. ‘i*n philosophy*, a hypothetical being that responds to stimuli as a normal person would but that does not experience consciousness’	0.3%	0.3%	0.3%	0%	0.4%	0.2%
9. Canadian ‘man conscripted for home defence’ (*Canadian Military slang* 1939–45)	0%	0%	0%	0%	0%	0%
10. ‘a computer controlled by another person without the owner's knowledge and used for sending spam or other illegal or illicit activities’	0.9%	0.9%	0.9%	0%	0%	0%

Appendix B – Typicality ratings (Study 2A)

Study 2A used a typicality rating task to identify several features stereotypically associated with ‘zombie’.

Participants: As in each study in this paper, participants were recruited through advertising on Google for a free personality test, which was administered after the main tasks.[4] Participants were restricted to native English-speakers raised in North America, 16 years of age

[4] Such ‘push strategies’ (recruiting participants not looking to participate in research) ensure participants are more ‘experimentally naïve’ and less motivated to provide what they think are experimenters’ intended responses (Haug, 2018). Samples collected with the present strategy have been previously compared against samples collected with other methods in replication studies. The present strategy has been consistently found to generate a diverse sample in terms of geography, socio-economic status, religiosity, political orientation, age, and education (e.g., Livengood et al., 2010; Sytsma, 2010; 2012; Sytsma & Ozdemir, 2019).

or older, with at most minimal training in philosophy.[5] 108 participants met these restrictions, with 26.9% of these failing the attention check. This left 79 participants.[6]

Methods: Each participant read 32 sentences attributing a feature to zombies (see Table B) and rated how typical each of these features are for zombies, on a 1-7 scale anchored at 1 with 'very atypical', at 4 with 'neutral (neither typical nor atypical)', and at 7 with 'very typical'. Material development was supported by a norming study, where sixteen psychology undergraduate students, all native speakers of English, first listed typical properties of zombies and then rated the typicality of features the experimenters had suggested based on dictionary explanations and their own intuitions. Based on the norming study, we included 18 presumptively typical zombie features and 14 presumptively atypical features (including two diagnostic of voodoo sense zombies). An attention check was included with the items: 'Please select 5 for this item'.[7] All items appeared in random order.

Results and discussion: To test the difference between the 18 tested features expected to be typical of zombies and the 14 features expected to be atypical of zombies, we began by running a repeated-measures ANOVA with *typicality* as a within-subjects factor, controlling for variation between participants across the 32 items. There was a significant effect for *typicality* and the effect size was large $F(1,2448)=1424$, $p<.001$, $\eta^2=.35$. As seen in Table B, the mean for each of the typical features was numerically above the mid-point, while the mean for each of the atypical features was below mid-point. Follow-up tests revealed that this difference was significant for 17 of 18 typical features and for 13 of 14 atypical features. Details for these tests are given in Table B.[8]

Further analyses were conducted to facilitate the assessment of further inferences from feature attributions. Feature attributions work together in promoting further inferences. 'Attacks and eats humans' promotes inferences to 'lion' and 'shark' – and these animals enjoy conscious experience. But 'attacks and eats humans and has a rotting body' may promote different inferences. As a preparatory step for the study of further inferences from attributions of typical zombie features, we therefore need to identify clusters of features that will be coactivated by 'zombie' and will jointly support onward inferences. While rigorous study of coactivation requires priming studies (Lucas, 2000), we reasoned that features that individuals tend to treat as similarly typical or similarly atypical will receive equal amounts of default activation or inhibition from the noun 'zombie' and will, *mutatis mutandis*, influence onward inferences to the same extent.

[5] Minimal training in philosophy was taken to exclude philosophy majors, those who have completed a degree with a major in philosophy, and those who have taken graduate-level courses in philosophy.

[6] These participants were 78.4% women (1 non-binary), mean age 40.9 years (16-67 years).

[7] 26.9% of participants failed the attention check. All the studies reported in this paper used a similar attention check, with just the number participants were asked to select varying.

[8] Throughout we use Student's t-tests for one-sample and paired-sample comparisons and Welch's t-tests for independent-sample comparisons. Since most predictions are directional, we use one-tailed tests unless specified otherwise (like here). We report nonparametric tests—either Wilcoxon signed-rank tests, *W*, or Wilcoxon rank-sum tests, *V*—in parentheses. We also used these tests to confirm the parametric tests reported in the main paper.

Table B

Typical and atypical features from Study 2A in order of descending mean rating, showing mean (SD) after the item and one-tailed comparisons to mid-point below.

Typical Features	Atypical Features
Zombies have rotting bodies. 6.14 (1.53) $t(78)$=12.47, p<.001, d=1.40 (V=2549.5, p<.001, r=0.79)	**Zombies are alert to their surroundings.** 3.90 (2.08) $t(78)$=.43, p=.33, d=.049 (V=831.5, p=.27, r=.036)
Zombies attack humans. 6.06 (1.65) $t(78)$=11.11, p<.001, d=1.25 (V=2670, p<.001, r=.75)	**Zombies feel thirsty.** 3.08 (1.75) $t(78)$=4.69, p<.001, d=.53 (V=214.5, p<.001, r=.46)
Zombies eat humans. 5.99 (1.67) $t(78)$=10.59, p<.001, d=1.19 (V=2476.5, p<.001, r=.74)	**Zombies are sad.** 2.86 (1.74) $t(78)$=5.83, p<.001, d=.66 (V=144, p<.001, r=.54)
Zombies move stiffly. 5.52 (1.92) $t(78)$=7.03, p<.001, d=.79 (V=2047, p<.001, r=.61)	**Zombies are under a magic spell.** 2.73 (1.84) $t(78)$=6.12, p<.001, d=.69 (V=179, p<.001, r=.56)
Zombies feel no joy. 5.52 (1.94) $t(78)$=6.96, p<.001, d=.78 (V=1969.5, p<.001, r=.62)	**Zombies feel hate.** 2.71 (1.63) $t(78)$=7.06, p<.001, d=.79 (V=110.5, p<.001, r=.62)
Zombies have been infected. 5.49 (2.01) $t(78)$=6.60, p<.001, d=.74 (V=2175, p<.001, r=.60)	**Zombies think.** 2.70 (1.67) $t(78)$=6.92, p<.001, d=.78 (V=133, p<.001, r=.61)
Zombies are dead inside. 5.49 (2.02) $t(78)$=6.58, p<.001, d=0.74 (V=1969.5, p<.001, r=.59)	**Zombies are under others' control.** 2.65 (1.71) $t(78)$=7.04, p<.001, d=.79 (V=159, p<.001, r=.62)
Zombies feel hungry. 5.42 (1.98) $t(78)$=6.35, p<.001, d=.71 (V=1905, p<.001, r=.58)	**Zombies are intelligent.** 2.61 (1.44) $t(78)$=8.57, p<.002, d=.96 (V=41, p<.001, r=.70)
Zombies have lifeless faces. 5.39 (1.91) $t(78)$=6.48, p<.001, d=.73 (V=2060.5, p<.001, r=.59)	**Zombies are alive.** 2.48 (1.69) $t(78)$=8.01, p<.001, d=.90 (V=97, p<.001, r=.66)
Zombies move slowly. 5.29 (1.85) $t(78)$=6.19, p<.001, d=.70 (V=1572.5, p<.001, r=.58)	**Zombies are happy.** 2.13 (1.59) $t(78)$=10.48, p<.001, d=1.18 (V=97, p<.001, r=.76)
Zombies have a rigid stare. 5.14 (2.00) $t(78)$=5.05, p<.001, d=.57 (V=1639, p<.001, r=.50)	**Zombies smell flowers.** 2.08 (1.53) $t(78)$=11.21, p<.001, d=1.26 (V=61, p<.001, r=.78)
Zombies lack free will. 4.95 (2.02) $t(78)$=4.18, p<.001, d=.47 (V=1456, p<.001, r=.42)	**Zombies talk.** 2.03 (1.54) $t(78)$=11.37, p<.001, d=1.28 (V=125, p<.001, r=.78)
Zombies feel no pain. 4.91 (2.06) $t(78)$=3.92, p<.001, d=.44 (V=1356, p<.001, r=.41)	**Zombies feel love.** 1.77 (1.27) $t(78)$=15.59, p<.001, d=1.75 (V=13, p<.001, r=.86)
Zombies smell blood. 4.91 (1.93) $t(78)$=4.20, p<.001, d=.47 (V=1315, p<.001, r=.45)	**Zombies sing.** 1.51 (0.96) $t(78)$=23.11, p<.001, d=2.60 (V=0, p<.001, r=.90)
Zombies have no feelings. 4.89 (2.17) $t(78)$=3.63, p<.001, d=.41 (V=1371.5, p<.001, r=.38)	
Zombies have been reanimated. 4.85 (1.84) $t(78)$=4.10, p<.001, d=.46 (V=1025.5, p<.001, r=.44)	
Zombies are dumb. 4.70 (1.60) $t(78)$=3.88, p<.001, d=0.44 (V=612, p<.001, r=.39)	
Zombies have no moods. 4.34 (2.21) $t(78)$=1.37, p=.087, d=.15 (V=1046, p=.063, r=.16)	

To identify clusters of features that are deemed similarly typical of zombies, we performed a hierarchical cluster analysis using the Ward algorithm with the Euclidian distance measure. Hierarchical clustering aims to group together items that are similar in the dimension measured by the data it is applied to (here: typicality ratings). The algorithm builds a hierarchy of clusters from the bottom up: each item starts in its own cluster, then the closest two are combined, with the process repeating until all items are in one cluster. To determine the similarity of single-item clusters, each participant's response for the item is treated as a feature of it, with these values locating the item in an n-dimensional space where n is the number of features (in this case 79, one for each participant in the study). Items that are closer together in this space are treated as more similar. As items are combined into multi-item clusters, the Ward algorithm combines them by minimizing the sum of squares. Applying this procedure for Study 2A produced the cluster dendrogram is shown in Figure B. In line with the previous analysis, we get two high-level clusters, with the 18 typical items clustering together on the right and the 14 atypical items clustering together on the left. We cut the dendrogram at height 18 to give six multi-item clusters – three clusters of typical items (T1-T3) and three clusters of atypical items (A1-A3). The least similar item from cluster A3 ("talk") was removed to reduce this cluster to four items, bringing it more in line with the others (2-3 items).

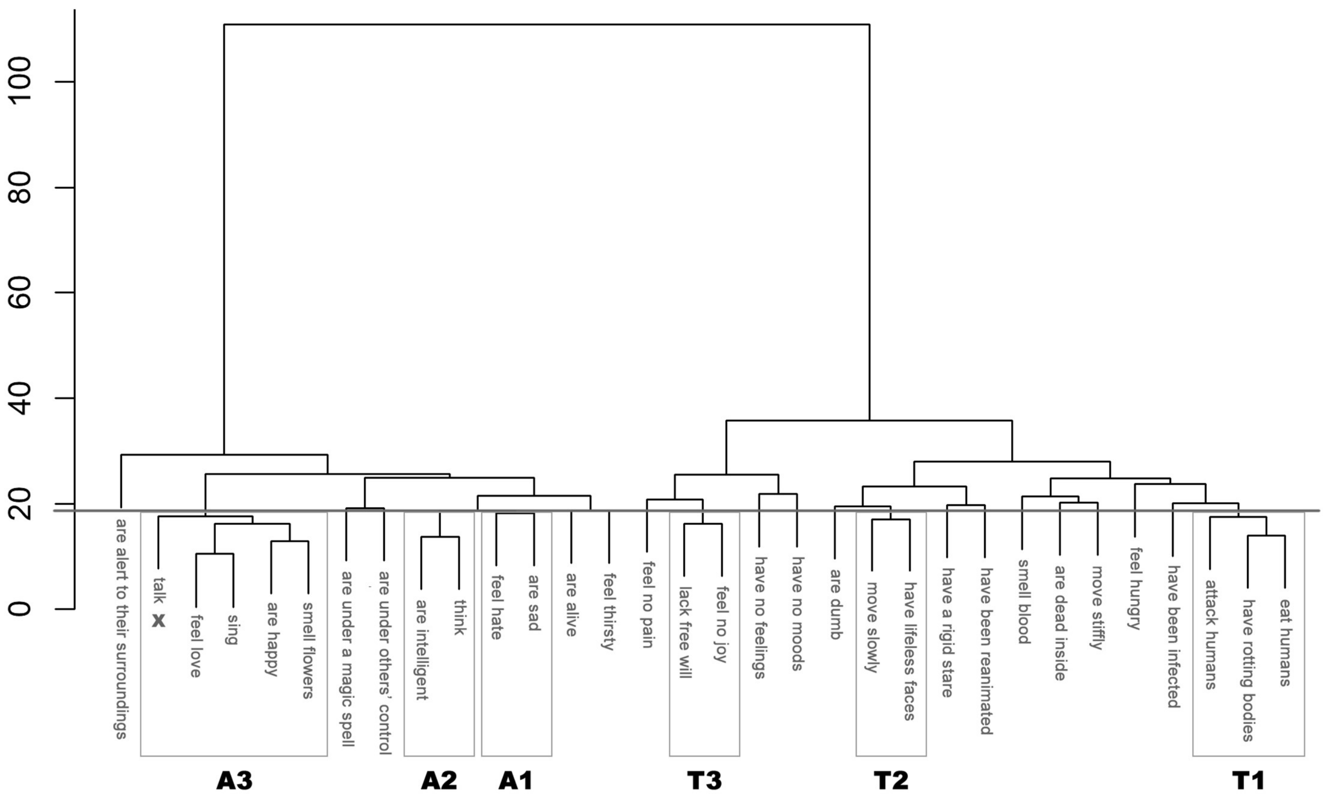

Figure B. Cluster dendrogram for Study 2A, showing cut and resulting clusters used for subsequent studies.

Appendix C – Plausibility ratings (Study 2B)

Study 2B employed a plausibility rating task to examine inferences from attributions of (clusters of) features that are, respectively, typical or atypical for Hollywood zombies, or typical for philosophical zombies. The study examined inferences to attributions of conscious experience (or its lack).

Participants: Participants were recruited in the same way, and with the same restrictions, as in Study 2A. 453 participants met these restrictions, with 34.7% failing one or both of the attention checks and a further 28.0% failing one or both of the comprehension checks described below. This left 213 participants.[9]

Methods: The study included two pages of critical items. On the first page, participants were invited to imagine 'biological beings' with different sets of properties. For each type of being, critical items then asserted that they enjoy conscious experience, using one of four different formulations: [10]

(A) these beings are capable of having conscious experiences;
(B) these beings have an inner mental life, including feelings and emotions;
(C) these beings are sentient and experience their surroundings and sensations;
(D) there is something it is like to be such a being.

In a between-subject design, each participant saw a version of the questionnaire with one of these four formulations. Participants were instructed: 'The following items invite you to imagine biological beings with certain properties. They then claim that [A/B/C/D]. How plausible is this claim in each case?' Participants were then asked to rate each of eight types of beings. Six were characterized as bearers of one of the six feature clusters T1-T3 and A1-A3 (see Figure B) The remaining two types were

(1) 'beings that have bodies like ours and behave like us' and
(2) 'beings that are alive but where everything is dark inside'.

For each of the eight types of being participants were given an item like (example for T1, version A):

> 'Imagine beings that have rotting bodies and attack and eat humans. These beings are capable of having conscious experiences.'

For each item, participants rated the plausibility of the claim on a 1-7 scale anchored at 1 with 'very implausible', at 4 with 'neutral (neither plausible nor implausible)', and at 7 with 'very plausible'. An attention check was included with the items on each of the two pages. All items appeared in random order.

The second page was designed to examine how all features activated by the noun 'zombie', taken together, influence attributions of consciousness. As before, participants were asked to rate the plausibility of consciousness attributions to different kinds of beings. This time, however, the beings were picked out by nouns: 'zombies', 'robots', 'humans', 'dolphins', 'geraniums', 'elves', and 'rocks'. For instance, for 'zombies' the item for consciousness attribution (A) read:

> 'Imagine Zombies. Zombies are capable of having conscious experiences.'

This page used the same 7-point scale as the previous page and included a similar attention check. Items again appeared in random order.

The seven nouns were chosen to give a range of contrasts, including their level of cognitive ability, whether they were fictional, and whether they were alive. 'Humans' was expected to anchor the high-end of the scale, while 'rocks' was expected to anchor the low end. These two items were used as engagement and comprehension checks: Participants who rated

---

[9] These participants were 70.9% women (3 non-binary), mean age 41.4 years (16-77 years).

[10] These formulations have been taken to entail or imply attributions of phenomenal consciousness (Chalmers, 1996, p.xi). Philosophers including Chalmers (2018) hold that this notion captures a folk-psychological concept. Empirical studies have provided a more nuanced picture (Peressini, 2014) and evidence to the contrary (Sytsma & Machery, 2010; Sytsma, 2012; 2016; Sytsma & Ozdemir, 2019). For present purposes we only assume items A-D capture features of mental lives that are relevant to the philosophical concept of conscious experience.

consciousness attributions to humans at or below mid-point or attributions to rocks at or above mid-point were excluded from further analysis, for any page. Upon exclusion, a similar number of participants remained in each condition: A (N=57), B (N=51), C (N=54), and D (N=51).

Results and discussion: Findings for the first page are shown in Figure C-1. To test whether the four formulations A-D elicit similar responses across items, we conducted a two-way mixed ANOVA, with *formulation* (A-D) as a between-subjects factor and *item* as a within-subjects factor. For the first page, the analysis revealed a main effect of *item* $F(7,1463)=114.22$, $p<.001$, $\eta^2=.27$, but no main effect of *formulation* $F(3,209)=1.56$, $p=.20$, $\eta^2=.005$, nor an interaction $F(21,1463)=1.24$, $p=.20$, $\eta^2=.009$. That is, ratings were not notably affected by the specific formulation used. We therefore combined conditions A-D for our analyses.

Follow-up tests for the first page indicate that participants tended to find it highly plausible to attribute consciousness (whatever way we phrased it) to (1) beings that 'have bodies like ours and behave like us', with ratings significantly different from mid-point $t(212)=14.42$, $p<.001$, $d=.99$, two-tailed ($V=16961$, $p<.001$, $r=.67$). Attributing consciousness to (2) beings that 'are alive but where all is dark inside' struck participants as less plausible $t(212)=5.75$, $p<.001$, $d=.39$, two-tailed ($V=7005.5$, $p<.001$, $r=.31$). Strikingly, however, participants still tended to rate (2) as distinctly plausible, as assessed against the mid-point $t(212)=6.23$, $p<.001$, $d=.43$, two-tailed ($V=12153$, $p<.001$, $r=.39$). We infer that, in the absence of informative context, 'all is dark inside' has an interpretation compatible with possession of conscious experience ('full of bad thoughts and feelings').

Consciousness attributions to bearers of our atypical zombie property clusters (A1-A3) were all deemed highly plausible, with ratings significantly above mid-point.[11] Interestingly, consciousness attributions to bearers of our typical zombie property clusters (T1-T3) varied notably in their plausibility. While ratings for each of the three clusters differed significantly from mid-point, the mean for T1 was *below* midpoint while the means for T2 and T3 were *above* mid-point.[12] In other words, participants tended to find attributions of conscious experience implausible for T1 but plausible for T2 and T3 (although significantly less plausible for T2 and T3 than for any of A1-A3).[13]

This finding is striking: The behavioural features included in T2 ('move slowly and have lifeless faces'), as well as the properties in T3 ('lack free will and feel no joy'), are intuitively suggestive of diminished conscious experience – and indeed, like T1, attract consciousness ratings that are lower than for beings that (1) 'have bodies like ours and behave like us'.[14] But, in T1, the component 'attack and eat humans' does *not* suggest diminished conscious experience and *cancels* the inference from the remaining component 'have rotting bodies' to the conclusion 'is dead', which would imply 'lacks conscious experience'. So precisely the feature cluster that is least suggestive of lack of conscious experience is the only one that supports inferences to this lack. This finding suggests that T1 attributions support these inferences because this feature cluster is diagnostic of zombies: Participants plausibly infer

[11] [A1] $t(212)=15.42$, $p<.001$, $d=1.06$ ($V=17254$, $p<.001$, $r=.72$); [A2] $t(212)=21.20$, $p<.001$, $d=1.45$ ($V=19072$, $p<.001$, $r=.79$); [A3] $t(212)=25.57$, $p<.001$, $d=1.75$ ($V=20363$, $p<.001$, $r=.84$).

[12] [T1] $t(212)=9.25$, $p<.001$, $d=.63$ ($V=3534.5$, $p<.001$, $r=.53$); [T2] $t(212)=4.63$, $p<.001$, $d=.32$, two-tailed ($V=11002$, $p<.001$, $r=.30$); [T3] $t(212)=3.88$, $p<.001$, $d=.27$, two-tailed ($V=11208$, $p<.001$, $r=.26$).

[13] [T2 vs A1] $t(212)=7.74$, $p<.001$, $d=.53$ ($V=1204.5$, $p<.001$, $r=.41$); [T2 vs A2] $t(212)=9.97$, $p<.001$, $d=.68$ ($V=914.5$, $p<.001$, $r=.49$); [T2 vs A3] $t(212)=10.25$, $p<.001$, $d=.70$ ($V=975.5$, $p<.001$, $r=.50$); [T3 vs A1] $t(212)=7.73$, $p<.001$, $d=.53$ ($V=1520$, $p<.001$, $r=.41$); [T3 vs A2] $t(212)=10.41$, $p<.001$, $d=.69$ ($V=959$, $p<.001$, $r=.49$); [T3 vs A3] $t(212)=11.1$, $p<.001$, $d=.76$ ($V=670$, $p<.001$, $r=.53$).

[14] [T1] $t(212)=17.36$, $p<.001$, $d=1.19$ ($V=15123$, $p<.001$, $r=.62$); [T2] $t(212)=6.71$, $p<.001$, $d=.46$, two-tailed ($V=7356.5$, $p<.001$, $r=.35$; [T3] $t(212)=7.34$, $p<.001$, $d=.50$, two-tailed ($V=8820.5$, $p<.001$, $r=.38$).

from T1 features that the beings in question are zombies and infer from this description that they lack conscious experience. The findings from page two speak directly to this suggestion.

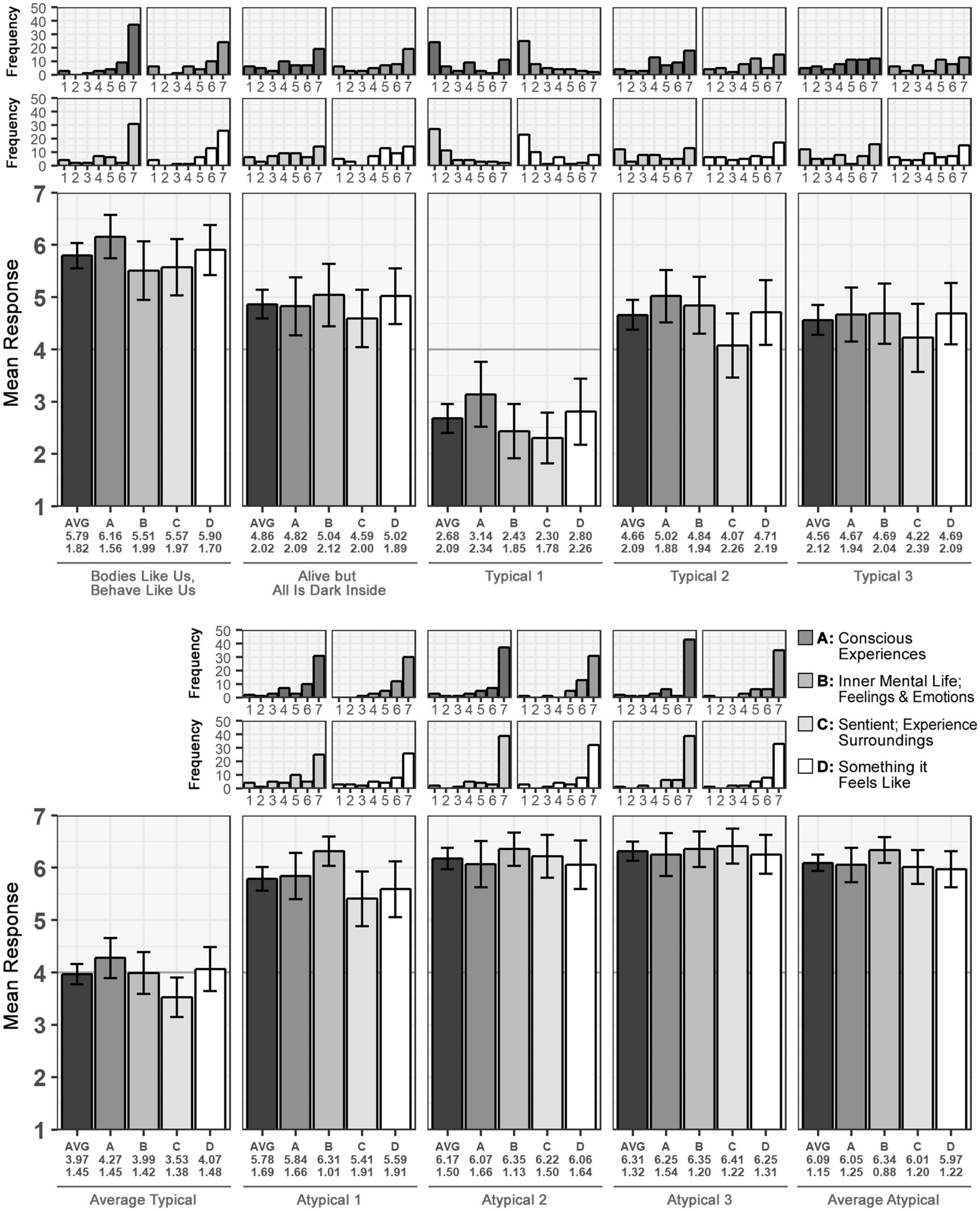

Figure C-1: Results for Study 2-B, Page 1 with means followed by standard deviations below the bar graphs; bar graphs showing 95% confidence intervals. Histograms above each bar graph show the frequency distributions of responses across participants for each condition.

Findings for the second page are presented in Figure C-2. A two-way mixed ANOVA for page two again revealed a main effect of *item* $F(4,836)=148.27$, $p<.001$, $\eta^2=.31$ and no main effect for *formulation* $F(3,209)=.51$, $p=.68$, $\eta^2=.001$, although there was now a significant interaction $F(12,836)=7.39$, $p<.001$, $\eta^2=.047$. Consciousness attributions to zombies, however, were similar across the four formulations, with a one-way ANOVA showing no effect for *formulation* $F(3,209)=1.00$, $p=.39$, $\eta^2=.014$. Combining formulations, consciousness attributions to zombies were deemed distinctly implausible (significantly below mid-point) $t(212)=15.25$, $p<.001$, $d=1.05$ ($V=1442$, $p<.001$, $r=.72$). Crucially, consciousness attributions to zombies attracted significantly lower ratings than consciousness attributions to bearers of any of the three clusters of typical zombie features (on page 1), including beings that (as per T1) 'have rotting bodies and attack and eat humans' $t(212)=3.16$, $p<.001$, $d=.22$ ($V=1982.5$, $p=.0014$, $r=.19$).

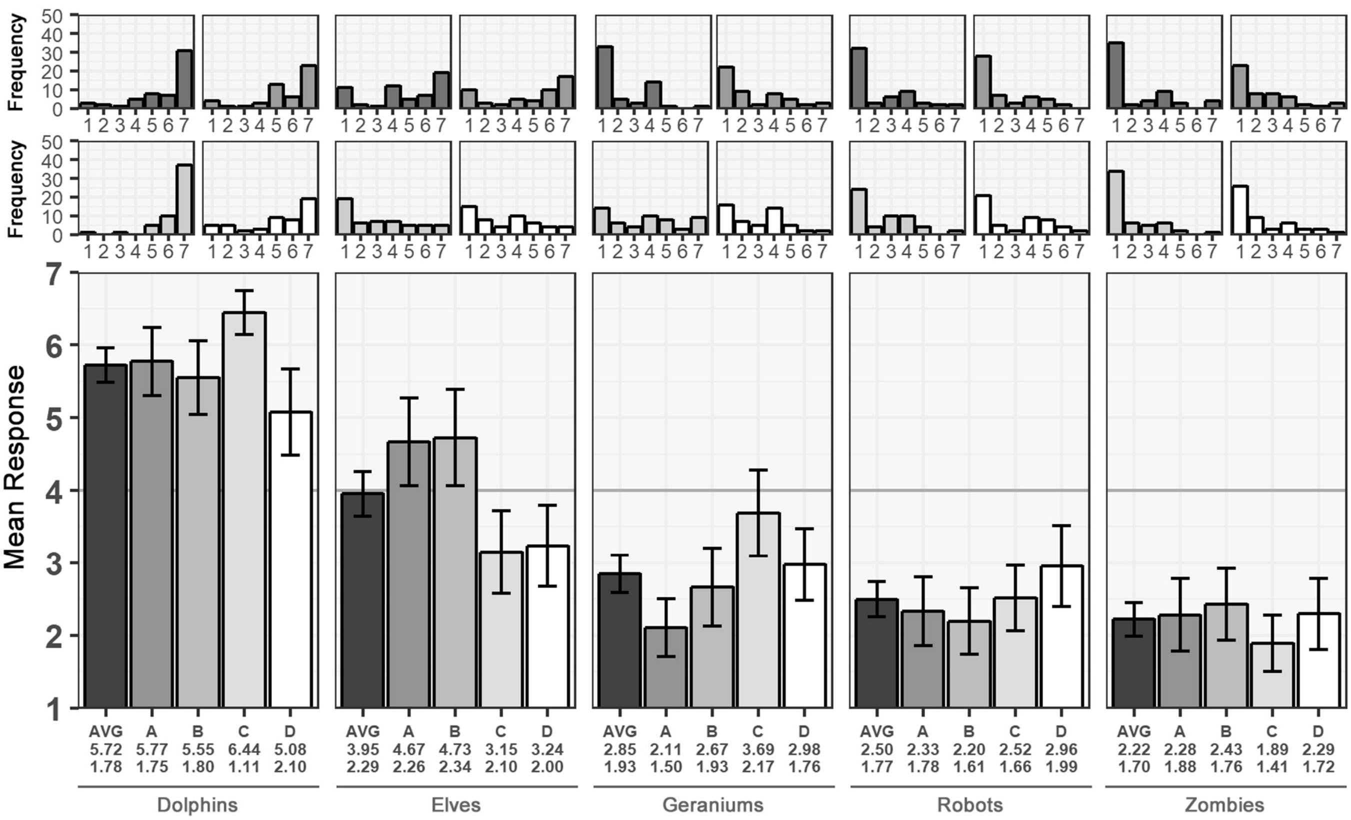

Figure C-2: Results for Study 2-B, Page 2 with means followed by standard deviations below the bar graphs; bar graphs showing 95% confidence intervals. Histograms above each bar graph show the frequency distributions of responses across participants for each condition.

Finally, our participants placed zombies on a very low rung of the ladder of consciousness visualized by Figure C-2, where zombies perch uneasily below robots and potted plants. Ratings for consciousness attributions to zombies were lower than for attributions to robots, for three of the four formulations (A, C, D), significantly so for two (C, D); there was no significant difference between the ratings for the fourth (B).[15] Similarly, ratings for zombies were lower than for geraniums for three of the four formulations (B, C, D), significantly so for two (C, D); there was no significant difference between the ratings for the fourth (A).[16]

Findings from this study suggest that outright lack of conscious experience is inferred not so much from (clusters of) typical features of zombies as from the noun itself. This noun triggers stereotypical inferences to *lacks conscious experience*. The stereotype associated with the dominant sense of 'zombie' includes the feature *lacks conscious experience* – even though this feature had not been produced in preparatory listing tasks (see Appendix B). Association with a negative feature (like *lacks conscious experience*) can be implemented through excitatory connections to a representation of this negative feature or inhibitory connections to its positive counterpart (*has conscious experience*) – or both. Excitatory connections predict typicality ratings above neutral for attributions of the negative feature, and inhibitory connections predict typicality ratings below neutral for attributions of the positive feature.[17] Studies 3A and 3B examine these predictions.

Appendix D – Typicality of possession of conscious experience (Study 3A)

Study 3A employed a typicality rating task (i) to examine whether the positive feature *possession of conscious experience* is atypical for zombies and (ii) to confirm whether clusters of individually typical features (like T1-T3) are also collectively typical of zombies.

Participants: 148 participants were recruited as before and met the basic restrictions, with 34.5% failing one or both of the attention checks. This left 97 participants.[18]

Methods: The study included two pages with critical items, with the pages counterbalanced for order.[19] On the first page, participants rated the previously used clusters of typical and atypical zombie features (T1-T3 and A1-A3) on the same scale as in Study 2A. Items appeared in random order and included an attention check. On the second page, participants rated how typical ten additional properties are for zombies, using the same scale. Items included the four attributions of consciousness to zombies used above (A-D). To contextualise their ratings, we further included three previously studied individual typical and atypical zombie features (*move stiffly*, *infected*, *feel hungry*, and *talk*, *alive*, *feel thirsty*, respectively). Items again appeared in random order and included an attention check.

Results and discussion: Findings are presented in Figure D. Starting with the first page, ratings for T1-T3 were all significantly above mid-point.[20] In contrast, ratings for A1-A3 were

---

[15] [A] $t(56)=.17$, $p=.43$, $d=.023$ ($V=182.5$, $p=.44$, $r=.045$); [B] $t(50)=1.02$, $p=.31$, $d=.14$, two-tailed ($V=192.5$, $p=.41$, $r=.048$); [C] $t(53)=2.51$, $p=.0076$, $d=.34$ ($V=85$, $p=.011$, $r=.15$); [D] $t(50)=1.99$, $p=.026$, $d=.28$ ($V=132.5$, $p=.020$, $r=.14$).

[16] [A] $t(56)=.64$, $p=.52$, $d=.085$, two-tailed ($V=196.5$, $p=.60$, $r=.031$); [B] $t(50)=.60$, $p=.28$, d=.084 ($V=279.5$, $p=.28$, $r=.063$); [C] $t(53)=5.76$, $p<.001$, $d=.78$ ($V=46$, $p<.001$, $r=.27$); [D] $t(50)=2.09$, $p=.021$, $d=.29$ ($V=181.5$, $p=.023$, $r=.13$).

[17] While we remain agnostic here about whether laypeople possess the concept of phenomenal consciousness, we assume they represent possession and lack of a *conscious experience* feature, on some understanding, and that it is entailed by all of formulations A-D above, on their intended interpretation. We further assume that inhibition of the *possession* representation and activation of the *lack* representation will influence typicality judgments on items like A-D in the same direction (if to slightly different extent, given evident differences between these items – e.g., between (B) feelings and (C) sentience).

[18] These participants were 74.2% women (1 non-binary), mean age 34.2 years (16-75 years).

[19] To check for order effects, we conducted a mixed ANOVA with *order* as a between-subjects factor and *item* as a within-subjects factor. No significant order effects were found.

[20] [T1] $t(96)=11.18$, $p<.001$, $d=1.14$ ($V=3738$, $p<.001$, $r=.72$); [T2] $t(96)=7.20$, $p<.001$, $d=.73$ ($V=2832$, $p<.001$, $r=.58$); [T3] $t(96)=4.10$, $p<.001$, $d=.42$ ($V=2445.5$, $p<.001$, $r=.38$).

all significantly below mid-point.[21] These findings confirm that these collections of individually typical zombie features are also collectively typical of zombies. Turning to the second page, typicality ratings for each of the three typical individual features were significantly above mid-point.[22] Ratings for each of the atypical individual features were significantly below mid-point.[23] The pattern observed replicated that observed for these features in Study 2A.

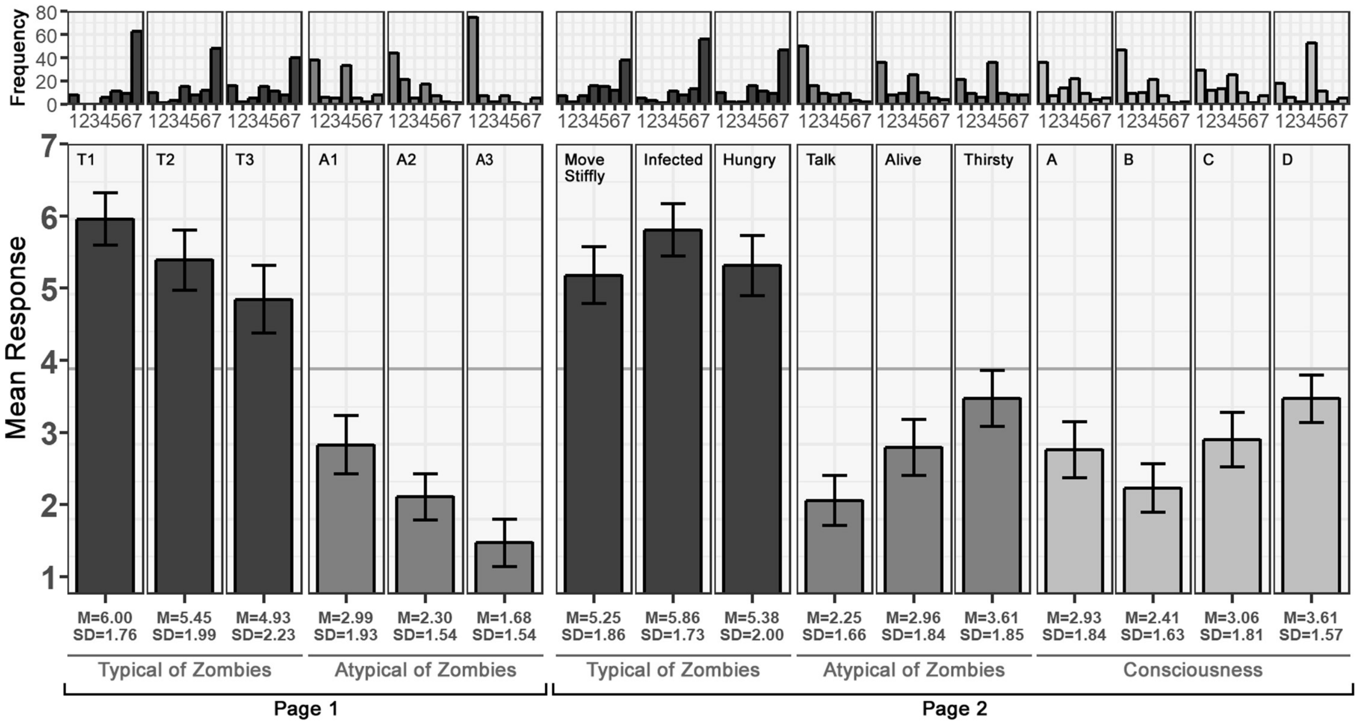

Figure D. Results for Study 3A with means followed by standard deviations below the bar graphs; bar graphs showing 95% confidence intervals. Histograms above each bar graph show the frequency distributions of responses across participants for each condition.

Turning to the crucial consciousness attributions (A-D), we examined variation between different formulations with a repeated-measures ANOVA. This showed a significant (if small) effect for *formulation* $F(3,288)=13.16$, $p<.001$, $\eta^2=.059$. Visual inspection suggests this was driven by somewhat lower ratings for B and somewhat higher ratings for D.[24] The difference is driven in part by the strikingly large proportion of participants answering ‘4’ (neutral) for formulation D (54.6%). This is notably higher than for the related attributions A (23.7%), B (21.6%), and C (25.8%). Since each of these formulations entails D on its intended Nagelian interpretation (e.g., if a being has an ‘inner mental life’, there is something it is like to be it), this suggests that well over a quarter of participants in this study did not understand D as

[21] [A1] $t(96)=5.16$, $p<.001$, $d=.52$ ($V=388.5$, $p<.001$, $r=.46$); [A2] $t(96)=10.86$, $p<.001$, $d=1.10$ ($V=151.5$, $p<.001$, $d=.74$); [A3] $t(96)=14.79$, $p<.001$, $d=1.50$ ($V=254.5$, $p<.001$, $r=.81$).

[22] [move stiffly] $t(96)=6.61$, $p<.001$, $d=.67$ ($V=2768.5$, $p<.001$, $r=.56$); [infected] $t(96)=10.55$, $p<.001$, $d=1.07$ ($V=3403.5$, $p<.001$, $r=.72$); [feel hungry] $t(96)=6.80$, $p<.001$, $d=.69$ ($V=2739$, $p<.001$, $r=.57$).

[23] [talk] $t(96)=10.41$, $p<.001$, $d=1.06$ ($V=296.5$, $p<.001$, $r=.73$); [alive] $t(96)=5.57$, $p<.001$, $d=.57$ ($V=440$, $p<.001$, $r=.48$); [feel thirsty] $t(96)=2.08$, $p=.020$, $d=.21$ ($V=640$, $p=.013$, $r=.19$).

[24] This was confirmed by a series of pairwise comparisons, which showed significant differences between each pair of items except A and C $t(96)=.75$, $p=.46$, $d=.076$, two-tailed ($V=622$, $p=.54$, $r=.062$). The difference is most pronounced between B and D $t(96)=5.75$, $p<.001$, $d=.58$, two-tailed ($V=278.5$, $p<.001$, $r=.49$).

intended or at all.[25]

Crucially, however, possession of conscious experience was deemed atypical of zombies on each formulation, with all ratings significantly below mid-point: [A] $t(96)=5.74$, $p<.001$, $d=.58$ ($V=499$, $p<.001$, $r=.50$); [B] $t(96)=9.62$, $p<.001$, $d=.98$ ($V=189.5$, $p<.001$, $r=.70$); [C] $t(96)=5.11$, $p<.001$, $d=.52$ ($V=531.5$, $p<.001$, $r=.46$); [D] $t(96)=2.46$, $p=.0079$, $d=.25$ ($V=277$, $p=.0048$, $r=.17$). To illustrate, being *capable of having conscious experiences* (as per A) was deemed as atypical of zombies as *being alive* $t(96)=.14$, $p=.89$, $d=.014$, two-tailed ($V=760$, $p=.94$, $r=.0081$). We infer that the 'zombie' stereotype involves inhibitory links to the component features of conscious experience.

Appendix E – Typicality of lack of conscious experience (Study 3B)

Study 3B examined the typicality of *lack of conscious experience* more directly.

Participants: 181 participants were recruited as before and met the restrictions, with 49.7% of these failing the attention check.[26] This left 91 participants.

Methods: They used the same 7-point scale to rate how typical ten properties are for zombies. The ten items included four attributions of *lack* of consciousness:

(A*) Zombies are incapable of having conscious experiences.
(B*) Zombies lack an inner mental life, for instance, they lack feelings and emotions.
(C*) Zombies are not sentient and do not experience sensations or their surroundings.
(D*) There is nothing it feels like to be a zombie.

We added the individual typical and atypical zombie features used in Study 3A. These had been chosen so as not to prime attributions of lack of consciousness. Items appeared in random order and included an attention check.

Results: Findings are presented in Figure E. Typicality ratings for the three typical features were again significantly above mid-point.[27] Ratings for the atypical features were again significantly below mid-point.[28] Patterns for both kinds of items replicated the ones observed in the previous studies. Combining ratings for these six items for Studies 2A, 3A, and 3B, a two-way mixed ANOVA, with *study* as a between-subjects factor and *item* as a within-subjects factor showed neither a significant main effect for *study* $F(2,264)=2.25$, $p=.11$, $\eta^2=.003$ nor a significant interaction effect $F(10,1320)=1.21$, $p=.28$, $\eta^2=.004$. These replications support the reliability of our typicality ratings.

Turning to the crucial consciousness attributions, a repeated-measures ANOVA showed a significant (if small) effect for *formulation* $F(3,270)=3.97$, $p=.009$, $\eta^2=.019$. Visual inspection suggested that D* is the outlier here. The Nagelian formulation again prompted the highest proportion of neutral '4' ratings (34.1%), suggesting many participants again failed to understand it as intended or at all. We therefore ran a follow-up ANOVA with D* removed.

[25] This peak at '4' was not so clearly observed in Study 2B which presented each participant with one of formulations A-D only. In the present study, juxtaposition of D with the arguably more readily intelligible formulations A-C may have made participants infer with the Maxim of Manner that D intended some different feature, which remained opaque to them, and rated the elusive feature 'neither typical nor atypical' because they felt unable to make any typicality (or other) judgment about it. This explanation suggests many participants in the main study felt they did not understand D at all.

[26] These participants were 64.8% women (1 non-binary), mean age 40.1 years (16-74 years).

[27] [move stiffly] $t(90)=5.56$, $p<.001$, $d=.58$ ($V=2226$, $p<.001$, $r=.51$); [infected] $t(90)=6.55$, $p<.001$, $d=.69$ ($V=2882.5$, $p<.001$, $r=.55$); [feel hungry] $t(90)=3.51$, $p<.001$, $d=.37$ ($V=1956$, $p<.001$, $r=.34$).

[28] [talk] $t(90)=11.41$, $p<.001$, $d=1.20$ ($V=113$, $p<.001$, $r=.76$); [alive] $t(90)=5.83$, $p<.001$, $d=.61$ ($V=462.5$, $p<.001$, $r=.52$); [feel thirsty] $t(90)=5.00$, $p<.001$, $d=.52$ ($V=333$, $p<.001$, $r=.46$).

This analysis did not show a significant effect for *formulation* $F(2,180)=2.28$, $p=.11$, $\eta^2=.009$. This allowed us to average across the items A*-C* that participants treated similarly. We thus found lack of conscious experience was deemed distinctly typical of zombies, with the mean rating significantly above the mid-point $t(90)=4.59$, $p<.001$, $d=.48$ ($V=2705.5$, $p<.001$, $r=.42$). Mean ratings for A*-C*, individually, were significantly above mid-point, and marginally so for D*.[29] To illustrate, being *incapable of having conscious experiences* (as per A*) was deemed as typical of zombies as *moving stiffly* $t(90)=1.38$, $p=.17$, $d=.15$, two-tailed ($V=573$, $p=.14$, $r=.15$). This supports the conclusion that *lack of conscious experience* is a component feature of the stereotype associated with the dominant sense of 'zombie'.

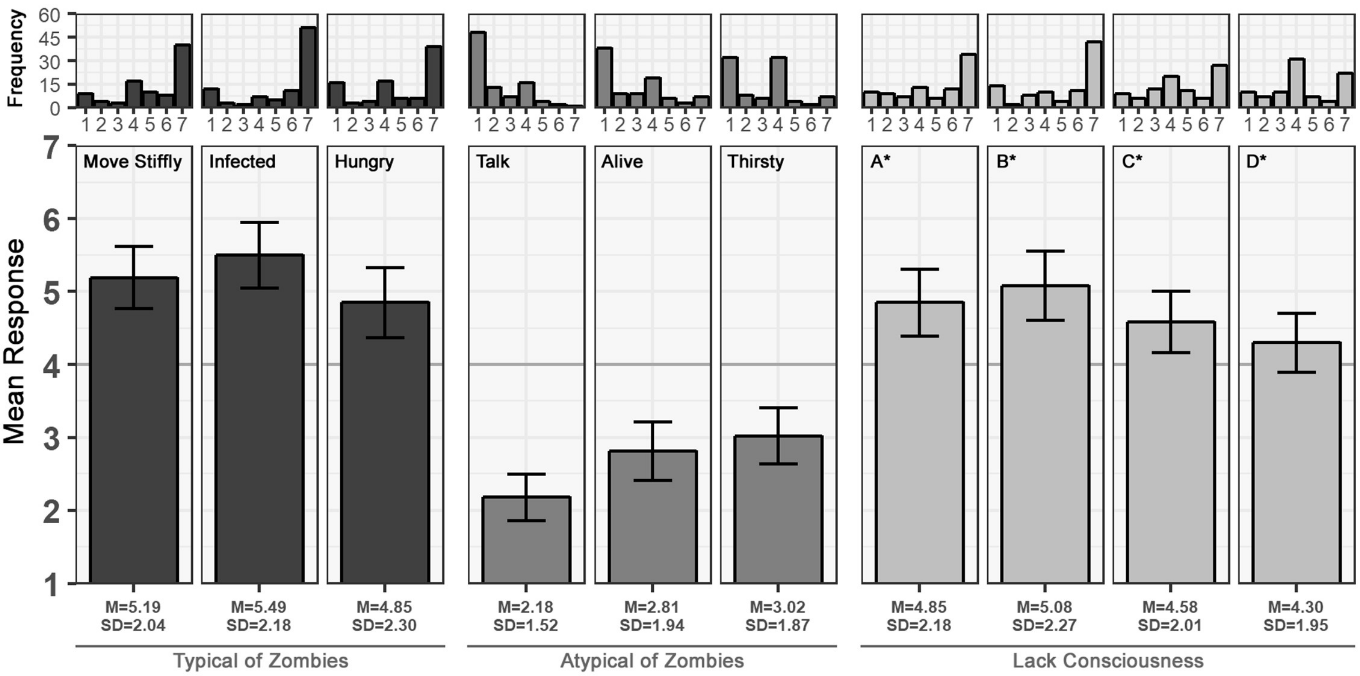

Figure E. Results for Study 3B with means followed by standard deviations below the bar graphs; bar graphs showing 95% confidence intervals, histograms shown above. Histograms above each bar graph show the frequency distributions of responses across participants.

Finally, the pattern of ratings for A*-D* inversely mirrors the pattern observed for positive attributions A-D in Study 3A (*cf.* Figures D and E). In fact, combining these items from Studies 3A and 3B, with A*-D* reverse coded, a two-way mixed ANOVA with *item* as a within-subjects factor and *study* as a between-subjects factor showed neither a significant main effect for *study* $F(1,186)=2.04$, $p=.16$, $\eta^2=.006$ nor a significant interaction $F(3,558)=.68$, $p=.56$, $\eta^2=.002$.

Appendix F – Conceivability judgments (Pre-study for Main Study)

Participants: 58 participants were recruited as before and using the same restrictions.[30]
Methods: Each participant read the following vignette describing the creation of a physical and

[29] [A*] $t(90)=3.70$, $p<.001$, $d=.39$ ($V=2251$, $p<.001$, $r=.37$); [B*] $t(90)=4.53$, $p<.001$, $d=.48$ ($V=2482$, $p<.001$, $r=.42$); [C*] $t(90)=2.76$, $p=.0035$, $d=.29$ ($V=1753.5$, $p=.0027$, $r=.27$); [D*] $t(90)=1.45$, $p=.075$, $d=.15$ ($V=1134$, $p=.050$, $r=.13$).

[30] These participants were 69.0% women, mean age 41.5 years (16-78 years).

behavioural duplicate (=**P**) that lacks conscious experience (=~**Q**):

> Here is a science-fiction story: In the future, scientists create humanoid beings. They scan the bodies of ordinary people, including their brains, at the molecular level. Using this information, the scientists then create an exact physical duplicate of a person's body, molecule by molecule. This creates a duplicate that has a body just like the original person and that behaves just like the original person. At the same time, all is dark inside for the duplicate. The duplicate lacks conscious experiences.

Participants were then asked whether they agreed or disagreed with each of two statements on a 7-point scale (anchored at 1 with 'totally disagree', at 4 with 'neither agree nor disagree', and at 7 with 'totally agree'):

> **[contradictory]** This story about duplicates is contradictory.
> **[conceivable]** It is conceivable that such duplicates might exist one day.

Results are presented in Figure F below.

The mean for [contradictory] was below the mid-point, though not significantly so $t(57)=1.47$, $p=.15$, $d=.19$, two-tailed ($V=232$, $p=.17$, $r=.20$). The mean for [conceivable] was significantly above the mid-point $t(57)=2.04$, $p=.046$, $d=.27$, two-tailed ($V=494.5$, $p=.069$, $r=.30$). The difference between the two ratings was significant $t(57)=2.43$, $p=.018$, $d=.46$, two-tailed ($V=182$, $p=.017$, $r=.31$). At first blush, this suggests that participants had a slight tendency to hold that the duplicates are not contradictory and are conceivable. However, we found only a negligible correlation between participant responses and this correlation was not significant $r=-0.040$, $p=.77$. Further, as seen in Figure D, there were a notable percentage of undecided '4' responses (39.7% for [contradictory], 34.5% for [conceivable]), including over one-quarter of participants answering '4' for both questions (25.9%). Finally, over one-quarter of participants gave problematic pairs of responses, either agreeing with both questions (17.2%) or disagreeing with both questions (8.6%).

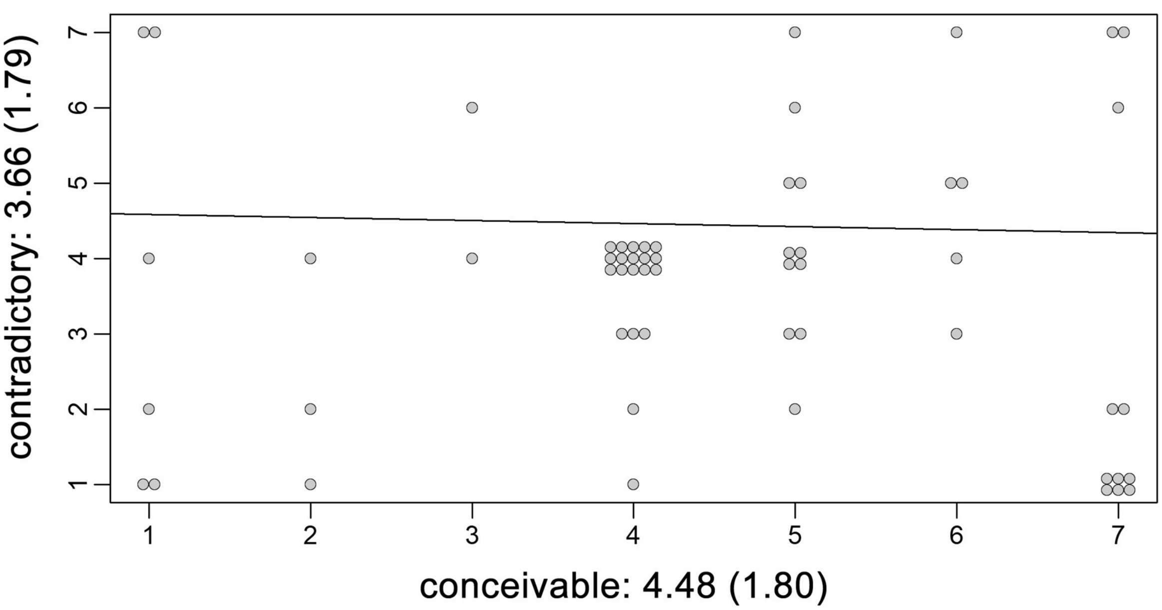

Figure F. Scatterplot for pre-study showing pattern of responses for [contradictory] and [conceivable], with linear regression line and responses spread out to show counts; mean next label with standard deviation in parentheses.

Appendix G – Effect of exclusions on effect of *term* and follow-up analysis (Main Study)

1. *Effect of Exclusions*

For the primary analyses for our main study that assessed **H1** and **H2**, we excluded participants who failed the attention check on the first page (‘Please select 5 for this item’), disagreed with the first comprehension on the second page (‘According to the story, a zombie has a brain just like the original person’s’), or agreed with the second comprehension check on the second page (‘According to the story, the zombie would behave differently than the original person).

To test the impact of these exclusions, we included all participants meeting the basic restrictions (native English-speakers raised in North America, 16 years of age or older, with at most minimal training in philosophy), whether or not they passed the attention and comprehension checks. We then repeated the mixed ANOVA reported in the main text, but adding two additional between-subjects factors: *attention* (passed / failed attention check on page 1) and *comprehension* (passed / failed comprehension checks on page 2). We again found a significant main effect of *term* $F(1,630)=32.35$, $p<.001$, $\eta^2=.019$; in addition, we found significant main effects for both *attention* $F(1,630)=9.69$, $p=.0019$, $\eta^2=.006$ and *comprehension* $F(1,630)=50.90$, $p<.001$, $\eta^2=.029$. No interaction effects were found between *term* and either *attention* $F(1,630)=0.47$, $p=0.49$, $\eta^2=.000$ or *comprehension* $F(1,630)=.021$, $p=.88$, $\eta^2=.000$, and similarly there were no higher-order interactions. This means that while participants who passed or failed the checks on each page did give significantly different responses, this did not significantly change the key effect of *term* that we are interested in.

To further check the exclusions, we calculated the mean response for each condition across all items (with T1-T3 reverse coded) for each of three groups of participants: [a] participants who failed the attention check, [b] participants who passed the attention check but failed the comprehension checks, and [c] participants who passed both the attention check and the comprehension check. We found that the difference in the means between the ‘duplicate’ and the ‘zombie’ conditions increased across these three groups: [a] 0.54 (4.38 for ‘duplicate’, 3.84 for ‘zombie’), [b] 0.58 (4.41, 3.83), [c] 0.66 (5.08, 4.41). Further, while the difference was significant for each group, the effect size was larger for [c] than for either [a] or [b]: [a] $t(159.93)=2.71$, $p=.0037$, $d=.41$; [b] $t(206.84)=3.10$, $p=.0011$, $d=0.43$; [c] $t(243.88)=4.07$, $p<.001$, $d=0.52$. A similar pattern holds looking at just the crucial consciousness attributions A-C: [a] 0.65 (4.07, 3.42), [b] 0.64 (4.30, 3.66), [c] 0.82 (5.02, 4.20). And, again, while the difference was significant for each group, the effect size was larger for [c] than for either [a] or [b]: [a] $t(169.63)=2.42$, $p=.0082$, $d=.36$; [b] $t(206.92)=2.51$, $p=.0064$, $d=.35$; [c] $t(243.41)=3.92$, $p<.001$, $d=.50$. These findings speak against the worry that the key difference between the ‘zombie’ and ‘duplicate’ conditions is due to cursory reading. In fact, quite the opposite: the effect of *term* is stronger among those paying attention and showing greater comprehension.

Finally, to test whether participants were able to conceive of philosophical zombies, we used an even stricter restriction, removing participants who answered ‘4’ on either comprehension check. An ANOVA for the consciousness attributions A-D with *term* and *restriction* (passed/failed further restriction) as between-participant factors, and controlling for variation between participants across the items, did not show a significant main effect for *restriction* $F(1,243)=2.47$, $p=.12$, $\eta^2=.005$ or a significant interaction effect $F(1,243)=.026$, $p=.87$, $\eta^2=.000$. Similarly, excluding D there was still no main effect for *restriction* $F(1,243)=2.50$, $p=.11$, $\eta^2=.007$ and no interaction effect $F(1,243)=.008$, $p=.93$, $\eta^2=.000$. This shows that while the introduction of yet stricter restrictions to examine the conceivability question was conceptually called for (see main text, Section 2.3), it did not notably affect ratings.

2. *Follow-up analyses for H1*

Given the effect of *cluster* found in the analysis in Section 2.3, we examined each item category separately. For the typical and atypical items, we ran a two-way mixed ANOVA with *term* as a between-subjects factor and *item* as a within-subjects factor. For the typical items (T1-T3), we found a main effect of *term* $F(1{,}245)=8.478$, $p=.0039$, $\eta^2=.017$, a main effect of *item* $F(2{,}490)=62.970$, $p<.001$, $\eta^2=.101$, and a marginally significant interaction $F(2{,}490)=2.644$, $p=.072$, $\eta^2=.004$. As predicted, follow-up tests showed that agreement is significantly higher in the zombie condition than in the duplicate condition for T1 $t(223.16)=2.92$, $p=.0019$, $d=.37$ ($W=8950.5$, $p=.0033$, $r=.17$) and T3 $t(243.9)=2.96$, $p=.0017$, $d=.38$ ($W=9232$, $p=.0018$, $r=.19$), although the difference was not significant for T2 $t(244)=.70$, $p=.24$, $d=.09$ ($W=7923.5$, $p=.29$, $r=.03$). For the atypical items (A1-A3), we again found a main effect of *term* $F(1{,}245)=12.29$, $p<.001$, $\eta^2=.033$ and a main effect of *item* $F(2{,}490)=21.634$, $p<.001$, $\eta^2=.025$. As predicted, follow-up tests showed that agreement is significantly higher in the duplicate than in the zombie condition for all three items [A1] $t(240.98)=2.40$, $p=.0086$, $d=.31$ ($W=6369$, $p=.012$, $r=.15$); [A2] $t(237.31)=3.41$, $p<.001$, $d=.43$ ($W=5783.5$, $p<.001$, $r=.21$); [A3] $t(243.44)=3.07$, $p=.0012$, $d=.39$ ($W=5955$, $p=.0013$, $r=.19$). Follow-up analyses for the consciousness items are reported in the main paper (Section 2.3).

Appendix H – Psycholinguistic interpretation

The main study observed small framing effects for attributions of typical and atypical zombie features (T1-T3 and A1-A3), but a medium-sized framing effect for consciousness attributions. Since conscious experience is not deemed more atypical of zombies than the other atypical features examined (A1-A3) (Study 3A, see Appendix D) and lack of consciousness no more typical than other relevant typical features (Study 3B, see Appendix E), the difference between consciousness and other attributions is unlikely to be due to different strengths of relevant stereotypical associations. Instead, it will be due to different levels of contextual support. The stereotypical inferences from 'zombie' to T1-T3 and A1-A3 clash with contextual information (physico-behavioural indistinguishability 'P') – and receive no contextual support from other parts of our vignette. Hence these inferences are largely suppressed and engender only a small difference between zombie and duplicate conditions. By contrast, our vignette, like the zombie argument, contains not only contextual information ('P') that cancels stereotypical inferences from 'zombie' to lack of conscious experience but also information that supports them ('all is dark inside'). This mitigates their suppression in the zombie condition. In the duplicate condition, where no inferences from the noun suggest lack of conscious experience, the perceived conflict between the implication (from 'P') of possession of conscious experience and the suggestion (from 'all is dark inside') of its lack is more likely to be resolved in favour of the former, by (re-) interpreting the latter (as 'full of dark thoughts and feelings'). This interpretation leaves a tension with 'P' (the average person's brain and behaviour will typically suggest less dysphoria) but renders the phrase ('all dark inside') consistent with attributions of conscious experience (whose lack is perceived as even more strongly inconsistent with ordinary physico-behavioural repertoire). This leads to larger differences between zombie and duplicate conditions, namely, the observed medium-sized framing effect for consciousness attributions.

The difference between the small framing effects observed in this study (for T1-T3 and A1-A3) and the large effects observed in previous studies of salience bias (Fischer & Engelhardt, 2019; 2020) may have two complementary explanations. All these effects result from the fact that the contextually relevant component feature of the dominant stereotype continues to pass on lateral co-activation to further component features that are frequently coinstantiated but irrelevant to the interpretation of the given subordinate use and cancelled by contextual information (see Section 1.3). In contrast with the previous studies, the present study employed a philosophical notion with internal tensions, where the contextually relevant component feature of the dominant zombie stereotype (*lack of conscious experience*) was simultaneously contextually cancelled (by 'P'). The relevant feature retains enough activation to notably influence ratings of consciousness attributions (resulting in the medium-sized framing effect). But its partial suppression does not leave it sufficiently strongly activated to pass on enough activation to the other features (e.g., T1-T3), for lateral cross-activation to have more than a small effect on the other judgments we elicited. If this account is correct, the fact that this study failed to observe large framing effects is due to the peculiar tension built into the philosophical notion of 'zombie' we employed.

A second explanation traces the differences between previously and currently observed effect sizes precisely to the differences between high vs low frequency and verb vs noun: Less frequent use of a word forges weaker associations between component features of the stereotype associated with it, so that less activation is passed on from contextually relevant features (like *lacks conscious experience*) to other component features (like T1-T3). This explains why framing effects observed with high-frequency words are large and the effects presently observed for attributions like T1-T3 and A1-A3 are small, and smaller than effects observed for attributions of consciousness (which are influenced directly by activation of the contextually relevant component feature). The fact that the framing effect observed for consciousness attributions was less than large could be due to the fact that stereotypes associated with nouns (other than event nouns) play a less central role than situation schemas associated with verbs in the construction of the situation models that underpin judgments about the cases described (*cf.* Melinger & Mauner, 1999; Tanenhaus & Carlson, 1989). Accordingly, noun-associated stereotypes influence these judgments to a lesser extent. If this second explanation is correct, low-frequency nouns will generally give rise to salience bias only in an attenuated form. Further research is required to decide to what extent these two potentially complementary explanations apply – and how strongly linguistic salience bias can arise from low-frequency nouns, in less peculiar and non-philosophical cases.

References

Baroni, M., Dinu, G. & Kruszewski, G. (2014), Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 238–247.

Chalmers, D.J. (1996). *The Conscious Mind*. New York: Oxford University Press.

Chalmers, D.J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25 (9-10), 6–61

Chang, T.M. (1986). Semantic memory: Facts and models. *Psychological Bulletin*, 99, 199-220.

Fischer, E., & Engelhardt, P.E. (2019). Eyes as windows to minds: Psycholinguistics for experimental philosophy. In E. Fischer & M. Curtis (eds.), *Methodological Advances in Experimental Philosophy* (pp.43-100). London: Bloomsbury

Fischer, E. and Engelhardt, P.E. (2020). Lingering stereotypes: Salience bias in philosophical argument. *Mind and Language*, 35, 415-439.

Fischer, E., Engelhardt, P.E. & Sytsma, J. (2020). Inappropriate stereotypical inferences? An adversarial collaboration in experimental ordinary language philosophy. *Synthese*. 2020; 1-42. https://doi.org/10.1007/s11229-020-02708-x

Firth, J.R. (1957). A synopsis of linguistic theory 1930–55. In *Studies in Linguistic Analysis*. Oxford: Blackwell, pp. 1–32.

Haug, M. (2018). Fast, cheap, and unethical? The interplay of morality and methodology in crowdsourced survey research. *Review of Philosophy and Psychology*, 9(2): 363-379.

Livengood, J., Sytsma, J., Feltz, A., Scheines, R., & Machery, E. (2010). Philosophical temperament. *Philosophical Psychology*, *23*, 313–330.

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin and Review*, *7,* 618–630.

Melinger, A., & Mauner, G. (1999). When are implicit agents encoded? Evidence from cross-modal priming. *Brain and Language, 68,* 185-191.

Peressini, A. (2014). Blurring two conceptions of subjective experience: Folk versus philosophical phenomenality. *Philosophical Psychology*, 27(6), 862–889.

Sytsma, J. (2010). Dennett's theory of the folk theory of consciousness. *Journal of Consciousness Studies,* 17 (3-4), 107-130.

Sytsma, J. (2012). Revisiting the Valence Account. *Philosophical Topics*, 40(2), 179-198.

Sytsma, J. (2016). Attributions of consciousness. In J. Sytsma & W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Oxford: Blackwell.

Sytsma, J., Bluhm, R., Willemsen, P. & Reuter, K. (2019). Causal attributions and corpus analysis. In E. Fischer & M. Curtis (Eds.), *Methodological Advances in Experimental Philosophy* (pp. 209-238). London: Bloomsbury.

Sytsma, J. & Machery, E. (2010). Two conceptions of subjective experience, *Philosophical Studies*, 151 (2), 299–327.

Sytsma, J. & Ozdemir, E. (2019). No problem: Evidence that the concept of phenomenal consciousness is not widespread. *Journal of Consciousness Studies*, 26 (9-10), 241-256.

Tanenhaus, M.K., & Carlson, G.N. (1989). Lexical structure and language comprehension. In W. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 529-561). Cambridge, MA: MIT Press.