Supplementary Information

Contents

Pseudo-words

Table 1. F0 and duration times of each pseudo-word according to two different recordings of samples. The first one was used for associative and judgement tasks. The second one was used for the recognition test following the 1x1 protocol, and does not contain training trials; four pseudo-words were also added in this recognition task, which explains the presence of grey cells. PW stands for pseudo-words.

		Sample 1		Sample 2 Recognition test (1x1 part 2)	
		Duration (ms)	f0 (Hz)	Duration (ms)	f0 (Hz)
PW used in associative and judgement tasks	[abab]	554	114	439	119
	[apap]	700	127	519	125
	[ibib]	683	118	494	141
	[ipip]	518	120	582	174
	[asas]	661	125	584	133
	[atat]	651	127	531	120
	[isis]	658	123	620	192
	[itit]	642	128	590	147
	[akak]	664	120	491	158
	[anan]	617	117	491	116
	[inin]	674	125	566	123
	[ikik]	692	118	659	167
	[ibib]	683	118	494	141
	[idid]	551	117	522	127
	[igig]	736	118	564	132
	[ilil]	598	114	508	134
	[ipip]	518	120	582	174
	[imim]	641	125	669	139
	[ubub]	609	119	539	134
	[ugug]	610	114	640	135
	[ukuk]	555	126	742	237
	[umum]	646	123	517	131
	[upup]	698	160	604	173
PW used in the	[yzyz]	622	115		
	[usus]	613	123		
	[ypyp]	658	126		

training part	[adad]	611	115		
	[agag]	563	119		
	[udud]	651	119		
	[ifif]	675	128		
	[aʁaʁ]	637	119		
PW added in the recognition task	[afaf]			520	135
	[uʃuʃ]			567	233
	[yʁyʁ]			492	149
	[yvyv]			561	134
	mean	634	122	562	150
	SD (Pearson)	0,054	8.11	0.067	30.50

Repartition of pseudo-words

Table 2. Distribution of pseudo-words in accordance with protocols and versions. PW stands for pseudo-words, V1 for version 1, V2 for version 2.

		1x1		1x2		2x1		2x2	
		V1	V2	V1	V2	V1	V2	V1	V2
Size	[abab]	x		x	x	x		x	x
	[apap]		x	x	x		x	x	x
	[ibib]		x	x	x		x	x	x
	[ipip]	x		x	x	x		x	x
Class	[asas]	x		x	x	x		x	x
	[atat]		x	x	x		x	x	x
	[isis]		x	x	x		x	x	x
	[itit]	x		x	x	x		x	x
Repulsiveness	[akak]	x		x	x	x		x	x
	[anan]		x	x	x		x	x	x
	[inin]	x		x	x	x		x	x
	[ikik]		x	x	x		x	x	x
Dangerousness	[ikik]	x				x			
	[ibib]	x				x			
	[idid]		x	x	x		x	x	x
	[igig]		x		x		x		x
	[ilil]	x		x	x	x		x	x
	[ipip]		x				x		
	[imim]		x	x			x	x	
	[ubub]		x	x	x		x	x	x
	[ugug]	x		x	x	x		x	x
	[ukuk]		x	x	x		x	x	x
	[umum]	x		x	x	x		x	x
	[upup]	x		x	x	x		x	x
Number of PW	21	12	12	20	20	12	12	20	20

Contingency tables

Aggregation of the results within each protocol

Contrasts of size are used here for the sake of illustration. Four pseudo-words were used ([abab], [apap], [ibib], [ipip]). In 2x2 (Fig. 1), answers were combined according to the target phonetic contrast that was presented (for the voicing consonant contrast, [abab] and [ibib] were combined, as well as [apap] and [ipip]; for the [i]-[a] contrast, [abab] and [apap] were combined, as well as [ibib] and [ipip]).

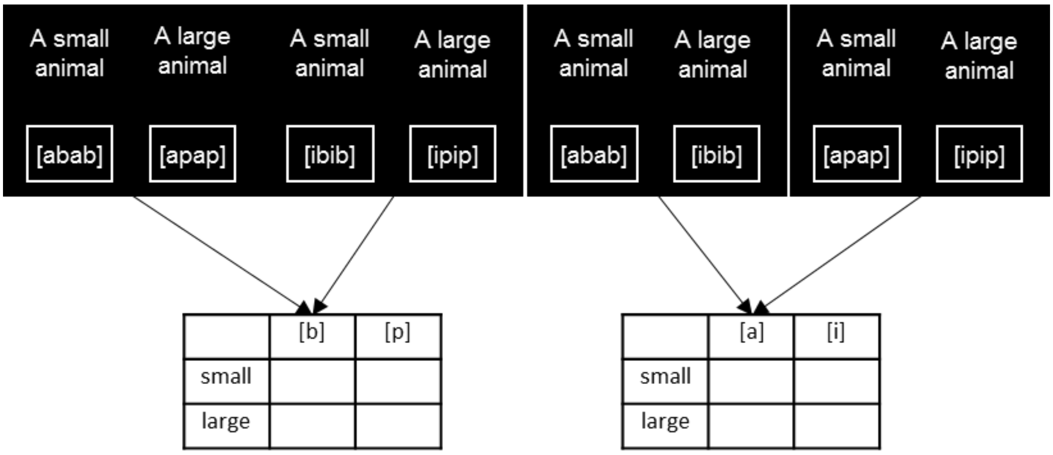

Fig. 1 Schematic representations of trials for the 2x2 protocol and resulting contingency tables. The two left representations contain consonantal contrasts; the two right ones contain vocalic contrasts.

In 1x2, the same combinations occurred but there were as many answers for 'small' as for 'large' (Fig. 2).

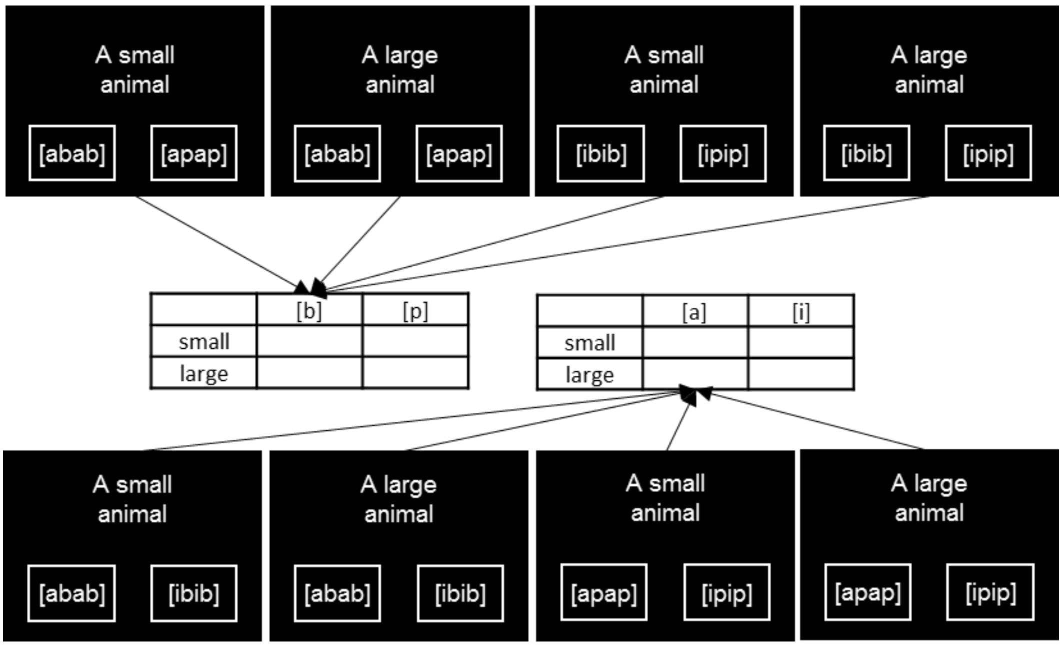

Fig. 2 Schematic representations of trials for the 1x2 protocol and resulting contingency tables. Representations at the top contain consonantal contrasts; the one at the bottom contain vocalic contrasts.

In 2x1, we simulated phonetic within-trial contrasts with contrasts between trials in order to produce comparable tables (Fig. 3). We obtained as many answers for the vowels as for the consonants.

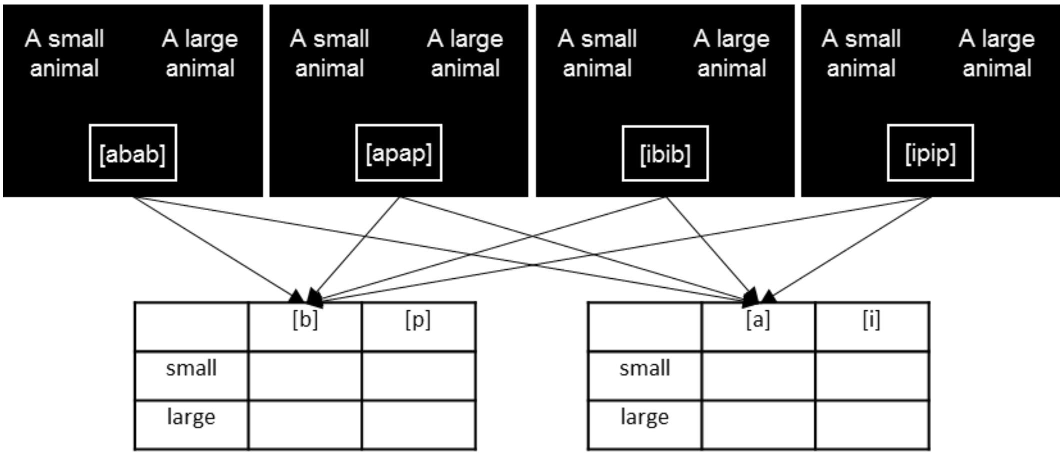

Fig. 3 Schematic representations of trials for the 2x1 protocol and resulting contingency tables. The output of each trial is added to the two contingency tables, according to the consonant and the vowel contained in the pseudo-word.

Finally, for the 1x1 protocol, there was the same number of answers for each combination (i.e. answers on a scale from 0 to 10), and the means of these answers across participants were computed (Fig. 4).

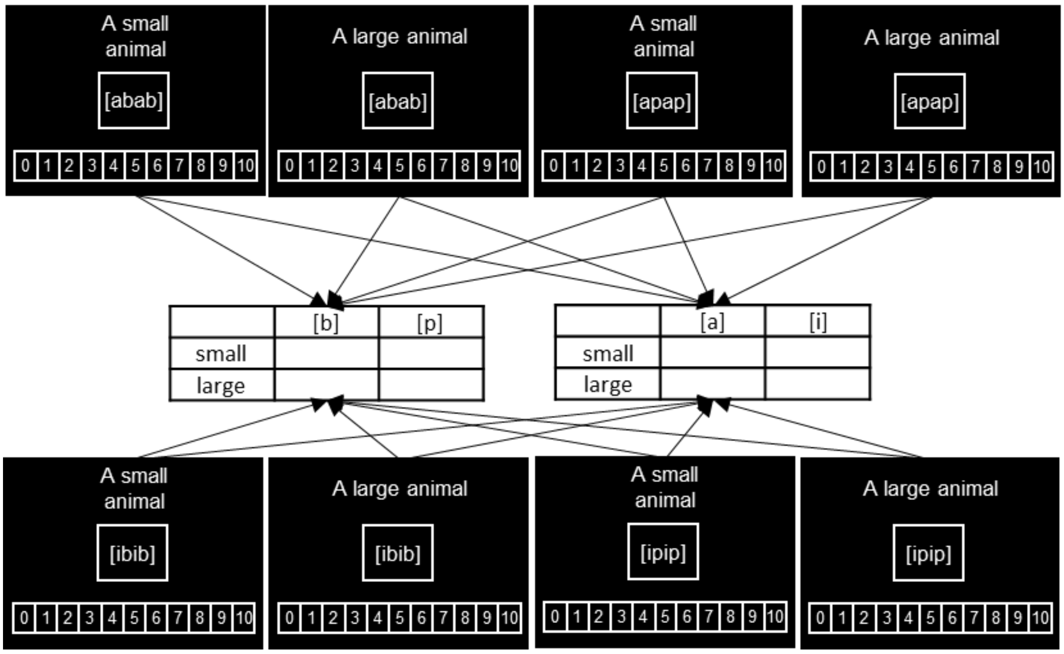

Fig. 4. Schematic representations of trials for the 1x1 protocol and resulting tables containing means. The result of each trial is added to the two contingency tables, according to the consonant and the vowel contained in the pseudo-word.

Steps in designing the analytical approach

Specifications of the hypotheses

On the one hand, given the specific instantiations of our general hypotheses, three of our conceptual categories – size, biological class and repulsiveness – are each associated with two target phonetic contrasts. One occurs between two consonants and the other between two vowels, which correspond to two hypotheses. The hypothesis on the consonantal contrast relies on the two vowels as possible contexts to create pseudo-words, while conversely the hypothesis on the vocalic contrast relies on the two consonants as possible contexts. For instance, for size, Hypothesis 1 opposes [p] to [b] with [i] and [a] as contexts, and Hypothesis 2 opposes [i] to [a] with [p] and [b] as contexts.

On the other hand, dangerousness is associated with four target phonetic contrasts:

- between [i] and [u], with [g] and [m] as contexts (Hypothesis 7 - contrast of vowels);
- between the two pairs of consonants {[b], [g]} and {[p], [k]}, with [u] as context (Hypothesis 8 - contrast of voicing);
- between the two pairs of consonants {[d], [b]} and {[l], [m]}, with [u] and [i] as contexts (Hypothesis 9 - contrast of manner);
- between [p] and [k], with [u] as context (Hypothesis 10 - contrast of place of articulation).

For contrasts of voicing or manner, pairs of consonants are opposed rather than single consonants and the place contrast which distinguishes the consonants in the pairs actually underlies the context rather than [u]. For Hypothesis 9, this consonantal context overlaps with the vocalic context: the target contrast of manner is presented either with alveolar consonants and the vowel [i] ([idid] and [ilil]), or with bilabial consonants and the vowel [u] ([ubub] and [umum]). Place only can thus be used when specifying the models. For Hypothesis 10, the target voicing contrast thus occurs in two contexts: bilabial consonants and [u], and velar consonants and [u], i.e. the place of articulation of the consonants is what matters.

Modelling subjects' judgments in 1x1

For the 1x1 protocol, different options seem available to relate the subjects' judgments to the features of the stimuli, but need to be considered carefully. Linear regression is not very well adapted, not mostly because of the integer values of the judgments, but because of floor and ceiling effects since these values cannot be lower than 0 and higher than 10. Poisson regression seems to address these issues in part, but scores on a Likert scale are not counts and are therefore unlikely to follow a Poisson distribution as required. Test and questionnaires can today be investigated with Item Response Theory-based methods (IRT), which can for instance appropriately account for differences in difficulty between items. In our case, with a single Likert scale repeated across trials, we can opt for a simpler approach and consider subjects' judgments as an ordered factor response and fit an ordered logistic regression model for each hypothesis.

Assessment of effect sizes

To assess the effect size of the key predictors of our various generalized linear regression models described in the previous section, we considered pseudo partial R^2.

In linear regression models, R^2, or coefficient of determination, is equal to the proportion of the variance in the dependent variable that can be explained by the predictors. It varies by definition between 0 and 1. By comparing the R^2 of a full model with all predictors with the R^2 of a model where one or some of the predictors have been dropped, it is possible to compute a coefficient of partial determination, or partial R^2. It represents the proportion of the variance in the dependent variable which can be explained by the predictor(s) in the full model. The partial R^2 for a predictor can be considered as a measure of the size of the effect of this predictor.

R^2 cannot be readily extended to generalized linear models, but many pseudo-R^2 resembling R^2 have been proposed by statisticians: McFadden pseudo-R^2 (McFadden, 1979), Cox & Snell pseudo-R^2 (Cox & Snell, 1989), Nagelkerke / Cragg & Uhler pseudo-R^2 (Nagelkerke, 1991), McKelvey & Zavoina pseudo-R^2 (McKelvey & Zavoina, 1975) etc. Several of these measures are derived from values of maximum likelihood computed for the target model and a corresponding baseline model. Experiments have been conducted to compare these different approaches and their specific perspectives on data (Smith & Mckenna, 2013; Veall & Zimmermann, 1996). Some authors have concluded that McKelvey & Zavoina pseudo-R^2 is the closest to least-squares R^2 in simple linear models, and offers the best comparability across various types of models (Veall & Zimmermann, 1996). We therefore chose this specific pseudo-R^2 to assess the effect sizes of our predictors across protocols.

Partial pseudo-R^2 can be computed in the same way partial R^2 are. We hence have a measure of effect size which can be applied to our different predictors and regression models across our four protocols. Partial McKelvey & Zavoina pseudo-R^2 is abbreviated as R^2_{MZ} in the main article and below.

As an assessment of R^2_{MZ} values as a relevant measure of effect size in generalized linear models, we compared it to better-known measures of effect size. For 2x2, 2x1 and 1x2, i.e. for contingency tables, we computed Cramer's V values, despite the non-independence of observation. For 1x1, we computed linear regression models in order to compute partial eta² (eta² are not available for a generalized linear models), despite the problematic distributions of residuals with deviance from normality and heteroscedasticity. As with the contingency tables for other protocols, we only considered the target predictor for the hypothesis at hand. For instance, for the first hypothesis, the predicted variable was ***judgment*** and the predictors were ***label***, ***vowel*** and their interaction (‘***judgment ~ label * V’***), and the effect size of the interaction was computed.

The point was here to verify the congruency between these different approaches. Fig. 5 and 6 report the relationships between the measures of effect size for the association tasks (2x2, 2x1 and 1x2 combined together) and for the judgment task (1x1), respectively.

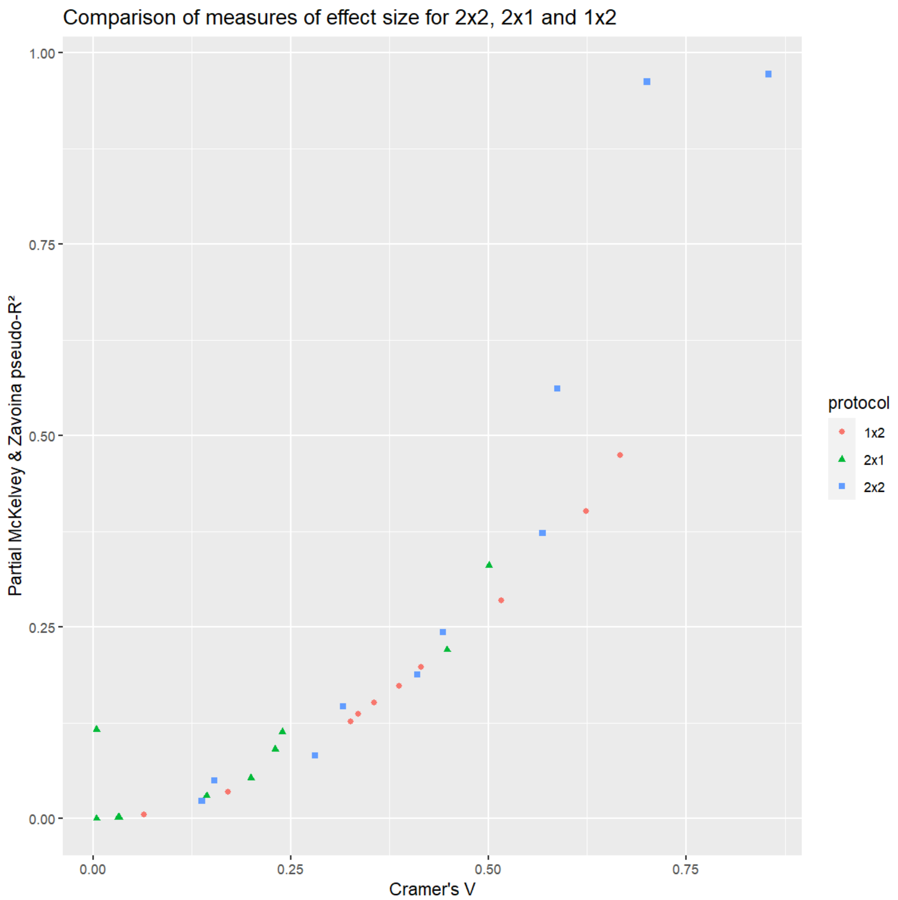

Fig. 5. Assessment of the correlation between two measures of effect size – R^2_{MZ} and Cramer's V – for the association tasks (2x2, 2x1 and 1x2).

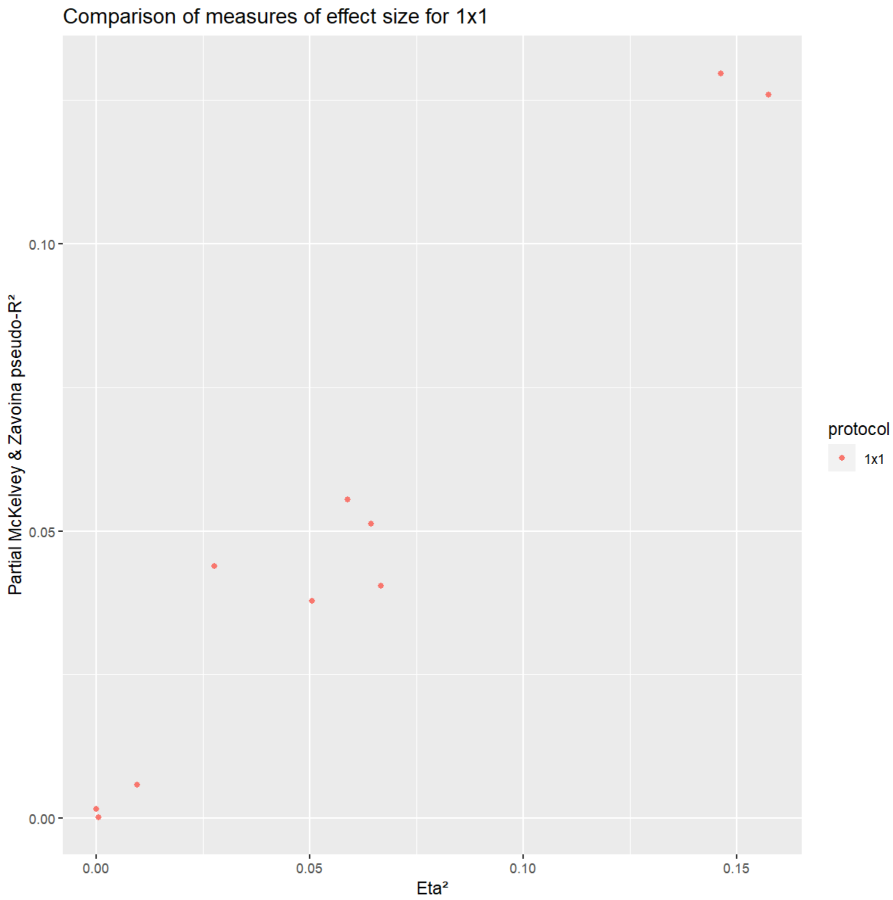

Fig. 6. Assessment of the correlation between two measures of effect size – R^2_{MZ} and eta² – for the judgment task (1x1).

It appears clearly that the different measures of effect size are congruent. Regarding Cramer's V for contingency tables and R^2_{MZ} for logistic regressions, the relationship seems more exponential than linear and Cramer's V values are higher than R^2_{MZ} values. Regarding eta² for linear regressions and R^2_{MZ} for ordered logistic regressions, the relationship is linear with very little to no difference in terms of magnitude of the two measures. The small divergences overall and possibly the lower values of R^2_{MZ} compared to Cramer's V are likely explained by the relative inadequacy of the simpler models.

These comparisons overall suggest that R^2_{MZ} is a reasonable measure of effect size, which can be applied to different generalized regression models and therefore across the four protocols.

Micro-analysis of data tables and models

In the 'Results by protocol and tested hypothesis' section of the article, statistically significant associations between linguistic and non-linguistic stimuli and their respective effect sizes were reported, but not the orientations of these associations, which are not revealed by partial R^2 and *p*-values. This section provides more details about sound symbolic associations as they appear by conceptual contrast across the different protocols.

The R^2_{MZ} and *p*-values presented in the following tables are the same as those presented in Tables 6 and 7 of the article. For 2x2, 1x2 and 2x1, these tables contain counts of response. For 1x1, they contain averaged judgments on a 0-to-10 scale. An additional statistical assessment of the various differences between pseudo-words could tell us more about the associations, e.g. is [i] associated more with 'small' than with 'large', or more with 'small' than [a]? However, this would lead to 240 (40 models x 6 assessments) tests, and finding then the right balance between preventing false positives (Type-I errors) and false negatives (Type-II errors) is difficult – see (Feise, 2002) for a discussion about this issue. For these reasons, we decided not to assess differences statistically, but to report propensities based on regressions that were significant and on what we could see in the contingency tables.

Simple effects for the significant interactions are also reported in this section. Holm-Bonferroni corrections would have led to less Type-I errors but increase the probability of Type-II errors. Since we judged more interesting here to risk overestimating effects than to potentially miss some, *p*-values have not been corrected.

Size (Hypotheses 1 & 2)

In 2x2, 1x2 and 2x1 for size contrasts, the effects of vowels and consonants are clear: there are associations between [a], [b] and 'large' and between [i], [p] and 'small' (Table 3). In 1x1, the effect of the consonants does not appear and the effect of vowels is weaker than those in other protocols. In 2x1 and 1x1, the patterns of response are similar: [p] and [i] are 'small'; [a] is more 'large' than 'small', but it is not as clear-cut as in the other protocols; [b] is neither 'large' nor 'small'. Overall, the sound symbolic associations tested for size are robust and thus well attested, with stronger effects for vowels than consonants.

Table 3. Contingency tables for contrasts of size in the four protocols.

	Vowels				Consonants			
		Large	**Small**			**Large**	**Small**	
2x2	**[a]**	9	6	***p* < .001**	**[b]**	12	8	***p* < .001**
	[i]	0	27	**R^2 = .962**	**[p]**	1	19	**R^2 = .561**

1x2	**[a]**	22	4	***p* < .001**	**[b]**	18	7	***p* = .015**
	[i]	7	25	**R^2 = .401**	**[p]**	14	23	**R^2 = .136**
2x1	**[a]**	30	17	***p* < .001**	**[b]**	24	23	***p* = .011**
	[i]	7	40	**R^2 = .331**	**[p]**	13	34	**R^2 = .113**
1x1	**[a]**	6.41	5.06	***p* = .006**	**[b]**	5.53	5.38	*p* = .141
	[i]	3.82	6.73	**R^2 = .130**	**[p]**	4.71	6.40	R^2 = .044

Biological class (Hypotheses 3 & 4)

The results about biological classes (Table 4) are difficult to interpret. There is an effect of vowels in 2x2 and 1x1, and an effect of consonants in 1x2. In 2x1 and 1x1, there is a bias in favor of 'bird': generally speaking, they are more often chosen or judged as more fitting with the presented pseudo-words. This bias cannot occur in 1x2, since there are as many answers for 'bird' as for 'fish', while it could have occurred in 2x2 and did not. This preferential bias may partly mask sound symbolic associations (with fewer answers for fish, less associations may be highlighted). However, the absence of a vocalic effect in 1x2 does not lend support to this idea. Moreover, 'fish' and 'bird' do not intrinsically oppose each other, and we could therefore have expected more sound symbolic associations in 2x2 and 2x1, where the conceptual contrast is explicit to subjects, than in 1x2 and 1x1. This was, however, not the case. Overall, contrary to the sound symbolic associations tested for size, those tested for bird vs. fish seem weak and are thus not well and comprehensively attested. Significant results, and beyond them all tables, nevertheless suggest that [i] and [t] associate with birds.

Table 4. Contingency tables for contrasts of biological class in the four protocols.

	Vowels				**Consonants**			
		Bird	**Fish**			**Bird**	**Fish**	
2x2	**[a]**	4	10	***p* = .021**	**[s]**	11	12	*p* = .252
	[i]	17	7	**R^2 = .188**	**[t]**	12	7	R^2 = .049
1x2	**[a]**	8	13	*p* = .238	**[s]**	13	23	***p* = .013**
	[i]	23	18	R^2 = .034	**[t]**	16	6	**R^2 = .151**
2x1	**[a]**	32	12	*p* = .964	**[s]**	32	13	*p* = .826
	[i]	34	13	R^2 < .001	**[t]**	34	12	R^2 = .002
1x1	**[a]**	4.76	3.82	***p* = .029**	**[s]**	5.82	3.63	*p* = .964
	[i]	7.44	3.81	**R^2 = .055**	**[t]**	6.44	4.00	R^2 < .001

Repulsiveness (Hypotheses 5 & 6)

For contrasts of repulsiveness, there is a clear effect of consonants, occurring in all protocols, irrespective of the within-trial (2x2, 1x2) or between-trial (2x1 and 1x1) condition (Table 5). Besides, a significant effect of vowels appears in 2x2 and 1x2, and there is a tendency for 2x1 and 1x1. There are two significant interactions (p = .037 for both) between the target contrast and the context for Hypotheses 5 and 6 in 2x1 (given the between-trial condition for phonetic contrasts and by construction of the stimuli, these two interactions are actually one and the same), and one (p = .011) for Hypothesis 6 in 1x2 (Fig. 7 & 8).

Fig. 7 shows that when presented with two labels (a repulsive animal and an attractive animal) in 2x1, the pseudo-word [akak] is much more strongly associated with repulsive animals than other pseudo-words. This explains the main effect that [a] associates with repulsive animals and [i] with attractive animals. Fig. 8 shows that in 1x2 pseudo-words with [i] are overall more attractive than those with [a], but much more when the contextual consonant is [k] than when it is [n]. On the one hand, that we do not see interactions in 2x2 may be explained by the masking of between-trial contrasts by within-trial contrasts. On the other hand, the lack of interaction in 1x1 may be due to the general weakness of this protocol to detect sound symbolic associations. No clear reason can be provided for why we don't see a second interaction when assessing consonants with 1x2. Overall, the sound symbolic association between repulsiveness and our consonants is well attested, while the association with our vowels is less robust across protocols. The significant interactions suggest specific word-level effects in addition to segment-level effects.

Table 5. Contingency tables for contrasts of repulsiveness in the four protocols. Interactions between the target phonetic contrast and the context are shown with a grey background.

	Vowels				Consonants			
		Attractive	**Repulsive**			**Attractive**	**Repulsive**	
2x2	**[a]**	2	20	**p < .001**	**[k]**	8	15	**p = .011**
	[i]	18	1	**R^2 = .972**	**[n]**	12	3	**R^2 = .243**
1x2	**[a]**	8	24	**p < .001**	**[k]**	9	20	**p = .018**
	[i]	20	6	**R^2 = .285**	**[n]**	21	12	**R^2 = .127**
2x1	**[a]**	14	32	p = .075	**[k]**	12	35	**p = .008**
	[i]	23	23	R^2 = .053	**[n]**	25	20	**R^2 = .116**
1x1	**[a]**	4.56	5.41	p = .069	**[k]**	4.33	6.06	**p = .004**
	[i]	6.22	4.75	R^2 = .038	**[n]**	6.44	4.06	**R^2 = .126**

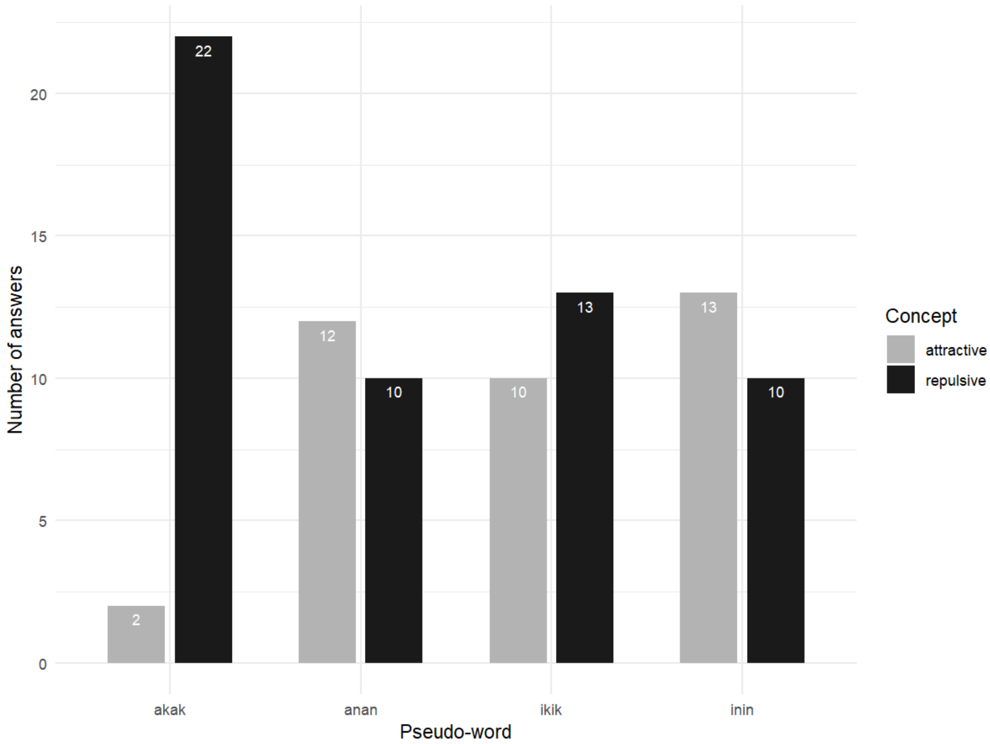

Fig. 7. Repartition of answers when assessing Hypotheses 5 and 6 (contrast of repulsiveness & contrasts of vowels and consonants) in 2x1.

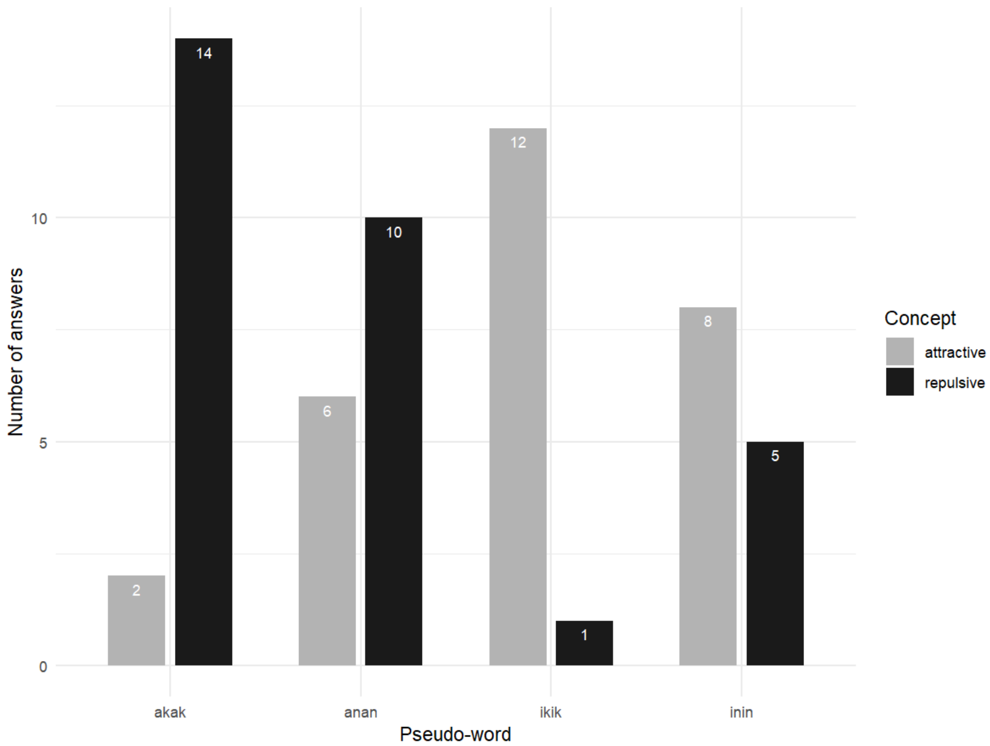

Fig. 8. Repartition of answers when assessing Hypothesis 6 (contrast of repulsiveness & contrast of vowel) in 1x2.

For the significant interaction in 2x1 (Hypotheses 5 & 6), the analysis of simple effects clearly shows that the pseudo-word [akak] differs from [anan], [ikik] and [inin], while other pairs of pseudo-words show similar behaviors in terms of preferences for repulsive or attractive (Table 6).

Table 6. Simple interaction effects found when assessing the vocalic and consonantal contrasts related to repulsiveness with the 2x1 protocol (Hypotheses 5 & 6). P-values smaller than 0.05 are in bold.

contrast	odds.ratio	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
k a / n a	0.076	0.065	Inf	0.014	0.404	-3.022	**0.003**
k a / k i	0.118	0.100	Inf	0.022	0.625	-2.513	**0.012**
k a / n i	0.070	0.059	Inf	0.013	0.370	-3.130	**0.002**
n a / k i	1.560	0.936	Inf	0.481	5.059	0.741	0.459
n a / n i	0.923	0.554	Inf	0.285	2.993	-0.133	0.894
k i / n i	0.592	0.352	Inf	0.184	1.899	-0.882	0.378

As for the significant interaction in 1x2 (Hypothesis 6), simple effects clearly show that the vocalic effect is stronger in the context of [k], compared to [n], since [akak] and [ikik] differ significantly, contrary to [anan] and [inin]. Since there is no significant difference between [akak] and [anan], nor between [inin] and [ikik], this vocalic contrast also likely underlies the significant difference between [akak] and [inin], and between [anan] and [ikik] (Table 7).

Table 7. Simple interaction effects found when assessing the [i-a] vocalic contrast related to repulsiveness with the 1x2 protocol (Hypothesis 6). P-values smaller than 0.05 are in bold.

contrast	odds.ratio	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
a k / i k	84.000	108.056	Inf	6.750	1045.307	3.444	**0.001**
a k / a n	4.200	3.845	Inf	0.698	25.264	1.568	0.117
a k / i n	11.200	10.604	Inf	1.751	71.637	2.552	**0.011**
i k / a n	0.050	0.058	Inf	0.005	0.488	-2.578	**0.010**
i k / i n	0.133	0.158	Inf	0.013	1.365	-1.698	0.090
a n / i n	2.667	2.051	Inf	0.591	12.042	1.275	0.202

Overall, all results are consistent in, ‘significantly’ or ‘near-significantly’, suggesting that [i] and [n] associate with attractive animals, and [a] and [k] with repulsive ones.

Dangerousness (Hypotheses 7 to 10)

Vowels (Hypothesis 7)

There is a significant effect of vowels in 1x2 and 2x1, and a tendency in 1x1, but no effect in 2x2, which is a bit surprising given the overall capacity of the latter protocol to reveal sound symbolic associations

(Table 8). There is also, quite surprisingly, a significant interaction (p = .002) between the target contrast and the context in 1x1 (Fig. 9), but not in the other protocols. This interaction corresponds to the specific pattern observed for the pseudo-word [ugug], which is judged both as more dangerous and less harmless (these two statements seem obviously complementary, but correspond to two judgments tasks) than the three other pseudo-words, which are judged in very similar ways by participants - the pattern is nearly opposite for [ugug]. The lack of significant effect for the vowels in 2x2 may be explained by an effect of the consonantal context (built with [g] or [m]). This is suggested by a significant effect of the consonantal context in 2x2 (p = .018 without correction, R^2_{MZ} = .159) and in 2x1 (p = .027 without correction, R^2_{MZ} = .082), and the pattern of interaction in 1x1. However, this runs rather opposite to our former idea that within-trial phonetic contrasts mask between-trial phonetic contrasts in 2x2 and 1x2 – in the present case, the within-trial vocalic contrast should mask the between-trial consonantal contrast. The situation is therefore unclear.

Table 8. Contingency tables for crossed contrasts of vowels and dangerousness in the four protocols. Interactions between the target phonetic contrast and the context are shown with a grey background.

		Vowels		
		Dangerous	**Harmless**	
2x2	**[i]**	5	15	p = .141
	[u]	11	10	R^2 = .082
1x2	**[i]**	5	17	***p* = .004**
	[u]	25	13	**R^2 = .198**
2x1	**[i]**	7	40	***p* = .036**
	[u]	16	30	**R^2 = .090**
1x1	**[i]**	3.87	7.37	p = .066
	[u]	5.22	5.79	R^2 = .051

Further assessing the significant triple interaction between target vowels, contextual consonants and labels in 1x1, judgments differ significantly for each pseudo-word according to the label it was presented with: the pseudo-words [igig] (p = .004), [imim] (p = .004) and [umum] (p = .001) are judged as significantly more fitting with 'harmless' rather than 'dangerous' animals; [ugug] is judged to fit more with 'dangerous'' rather than 'harmless' animals (p = .018) (Table 9). Hence, the combination between [g] and [u] induces stronger associative judgements with 'dangerous' animals, which differs from the other conditions. This being said, the global pattern of associations is difficult to interpret, and providing a full picture of it is beyond the target of this article.

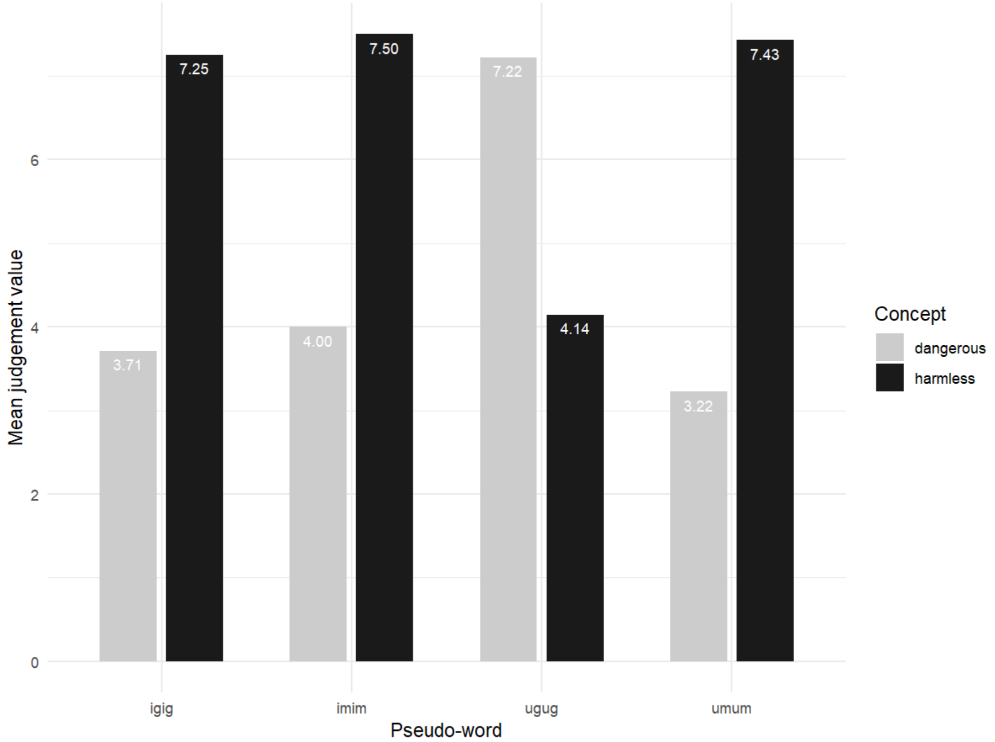

Fig. 9. Mean judgments when testing Hypothesis 7 (contrast of dangerousness and contrast of vowel frontness) in 1x1.

Table 9. Simple interaction effects in the 'triple' interaction found when assessing the [i-u] vocalic contrast for dangerousness with the 1x1 protocol (Hypothesis 7). P-values smaller than 0.05 are in bold.

contrast	estimate	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
dangerous i g - harmless i g	-2.772	0.962	Inf	-4.658	-0.887	-2.882	**0.004**
dangerous i g - dangerous u g	-2.512	0.904	Inf	-4.283	-0.741	-2.780	**0.005**
dangerous i g - harmless u g	-0.239	0.933	Inf	-2.067	1.589	-0.256	0.798
dangerous i g - dangerous i m	-0.175	0.840	Inf	-1.821	1.471	-0.209	0.835
dangerous i g - harmless i m	-2.836	0.939	Inf	-4.676	-0.996	-3.021	**0.003**
dangerous i g - dangerous u m	0.464	0.806	Inf	-1.117	2.044	0.575	0.565
dangerous i g - harmless u m	-3.016	1.040	Inf	-5.054	-0.978	-2.901	**0.004**
harmless i g - dangerous u g	0.260	0.886	Inf	-1.476	1.997	0.294	0.769
harmless i g - harmless u g	2.533	1.017	Inf	0.539	4.527	2.490	**0.013**
harmless i g - dangerous i m	2.597	0.951	Inf	0.733	4.460	2.731	**0.006**
harmless i g - harmless i m	-0.064	0.892	Inf	-1.812	1.684	-0.071	0.943
harmless i g - dangerous u m	3.236	0.951	Inf	1.372	5.100	3.402	**0.001**
harmless i g - harmless u m	-0.244	1.000	Inf	-2.204	1.716	-0.244	0.807
dangerous u g - harmless u g	2.273	0.961	Inf	0.389	4.157	2.364	**0.018**
dangerous u g - dangerous i m	2.336	0.890	Inf	0.592	4.081	2.625	**0.009**
dangerous u g - harmless i m	-0.324	0.856	Inf	-2.002	1.353	-0.379	0.705
dangerous u g - dangerous u m	2.975	0.891	Inf	1.228	4.723	3.338	**0.001**
dangerous u g - harmless u m	-0.504	0.974	Inf	-2.413	1.405	-0.518	0.605

contrast	estimate	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
harmless u g - dangerous i m	0.063	0.931	Inf	-1.761	1.887	0.068	0.946
harmless u g - harmless i m	-2.597	0.993	Inf	-4.544	-0.650	-2.614	**0.009**
harmless u g - dangerous u m	0.702	0.916	Inf	-1.093	2.498	0.767	0.443
harmless u g - harmless u m	-2.777	1.093	Inf	-4.919	-0.635	-2.542	**0.011**
dangerous i m - harmless i m	-2.660	0.926	Inf	-4.475	-0.846	-2.873	**0.004**
dangerous i m - dangerous u m	0.639	0.823	Inf	-0.974	2.252	0.776	0.438
dangerous i m - harmless u m	-2.841	1.030	Inf	-4.860	-0.822	-2.757	**0.006**
harmless i m - dangerous u m	3.299	0.927	Inf	1.482	5.117	3.557	**0.000**
harmless i m - harmless u m	-0.180	0.978	Inf	-2.096	1.736	-0.184	0.854
dangerous u m - harmless u m	-3.480	1.030	Inf	-5.498	-1.461	-3.379	**0.001**

Overall, all results are consistent in, ‘significantly’ or ‘near-significantly’, suggesting that [i] associates with harmless animals.

Voicing (Hypothesis 8)

There is no significant effect of the voicing contrast once *p*-values have been corrected, but near significance for 2x2 and 1x1 (uncorrected *p*-values are below 0.05) (Table 10). This pattern of near-significance is surprising given general observations about the ability of protocols to reveal sound symbolic associations. A significant interaction (p = .044) between the target phonetic contrast and the context appears in 2x1. Fig. 10 shows opposite patterns for [ukuk] and [upup], which elicit more extreme and opposite reactions than [ubub] and [ugug]. The effect of voicing, if there is one, may be masked by an effect of place, which could in turn explain the lack of significance in other protocols.

Table 10. Contingency tables for crossed contrasts of voicing and dangerousness in the four protocols. Interactions between the target phonetic contrast and the context are shown with a grey background.

Voicing

		Dangerous	Harmless	
2x2	**[b] & [g]**	13	3	p = .058
	[p] & [k]	12	12	R^2 = .146
1x2	**[b] & [g]**	18	16	p = .697
	[p] & [k]	13	15	R^2 = .005
2x1	**[b] & [g]**	21	25	p = .826
	[p] & [k]	23	24	R^2 = .001
1x1	**[b] & [g]**	5.94	5.31	p = .069
	[p] & [k]	3.59	6.25	R^2 = .040

Further assessing the simple effects of the significant interaction between voicing and place in 2x1, a single difference is found to be significant, between voiced posterior and voiceless posterior, i.e. between [ukuk] and [upup] (Table 11). No other significant difference between pseudo-words is found, which means that only an effect of place in an unvoiced context can be detected.

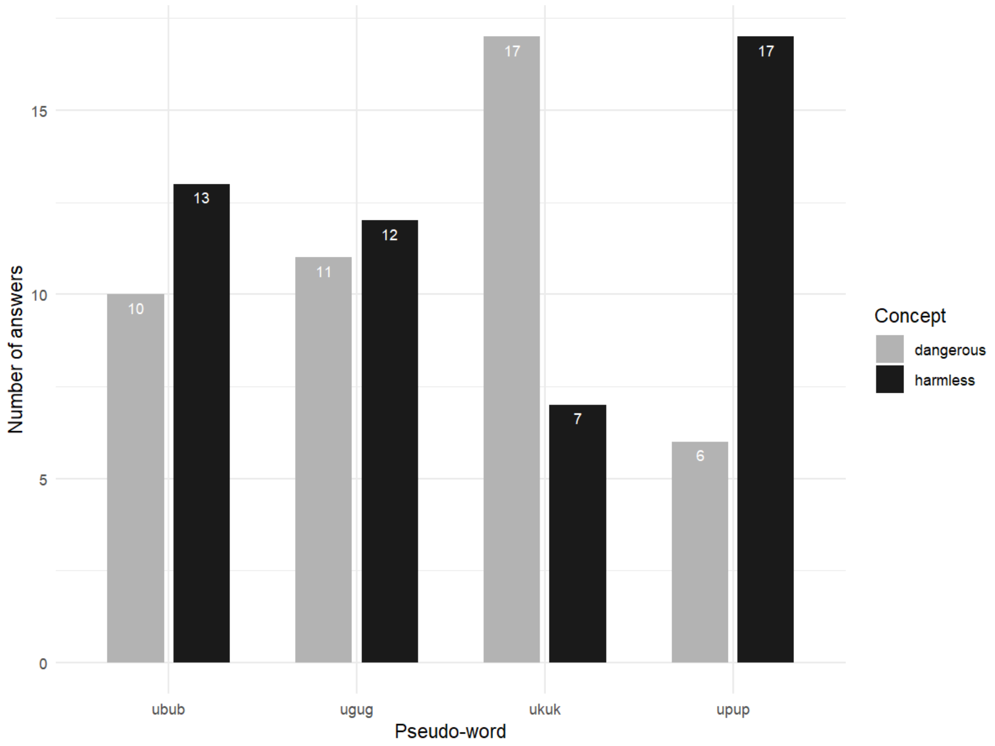

Fig. 10. Repartition of answers when assessing Hypothesis 8 (crossed contrasts of repulsiveness & voicing) in 2x1

Table 11. Simple interaction effects found when assessing the voicing contrast related to repulsiveness with the 2x1 protocol (Hypothesis 8). P-values smaller than 0.05 are in bold.

contrast	odds.ratio	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
voiced anterior / voiceless anterior	0.459	0.291	Inf	0.132	1.591	-1.228	0.219
voiced anterior / voiced posterior	1.192	0.706	Inf	0.373	3.807	0.296	0.767
voiced anterior / voiceless posterior	3.157	1.943	Inf	0.945	10.545	1.868	0.062
voiceless anterior / voiced posterior	2.597	1.642	Inf	0.752	8.968	1.510	0.131
voiceless anterior / voiceless posterior	6.881	4.497	Inf	1.911	24.773	2.951	**0.003**
voiced posterior / voiceless posterior	2.649	1.624	Inf	0.797	8.811	1.589	0.112

There is weak suggestion overall that the voiced consonants [b] and [g] associate with dangerous animals.

Manner (Hypothesis 9)

The effect of the consonant is only significant in 2x2 and 1x2 (Table 12). There seems to be preferential biases for ‘harmless’ animals in 2x2, 2x1 and 1x1, just as there was a preferential bias for ‘bird’ when investigating biological class. The lack of within-trial phonetic contrast may explain the lack of effect in 2x1 and 1x1. Manner therefore seems to interact only moderately with dangerousness, with the following orientation: plosive consonants with dangerous animals, and sonorants with harmless ones.

Table 12. Contingency tables for crossed contrasts of manner and dangerousness in the four protocols.

		Manner		
		Dangerous	**Harmless**	
2x2	**[b] & [d]**	17	10	***p* < .001**
	[m] & [l]	3	30	**R^2 = .373**
1x2	**[b] & [d]**	24	10	***p* = .003**
	[m] & [l]	13	28	**R^2 = .173**
2x1	**[b] & [d]**	15	29	*p* = .238
	[m] & [l]	10	37	R^2 = .030
1x1	**[b] & [d]**	4.00	6.53	*p* = .496
	[m] & [l]	2.89	6.56	R^2 = .006

Place (Hypothesis 10)

There are significant effects of the place of articulation in 2x1 and 1x2 (Table 13). The fact that it does not appear in 2x2 may be due to the small number of trials (only half of the participants where shown the contrast [upup]-[ukuk] – the other half was shown [ukuk] in a voicing contrast with [ugug]). Judgments for [ukuk] with ‘harmless’ animals in 1x1 are on average quite high (7.22), which is surprising at first since it is not strongly associated with ‘harmless’ in 2x1 and in 2x2 (15% and 24% of the associations made respectively, 25% being chance level. However, apart from being a back consonant, [k] is also voiceless and voiceless consonants ([p] and [k] in our study) are judged as better suited to ‘harmless’ animals (6.25) than to ‘dangerous’ ones (3.59) in voicing contrasts in 1x1. Overall, place therefore seems to interact only moderately with dangerousness, with the following orientation: the front consonant [p] with harmless animals, and the back consonant [k] with dangerous ones.

Table 13. Contingency tables for crossed contrasts of place and dangerousness in the four protocols.

		Place		
		Dangerous	**Harmless**	
2x2	**[p]**	4	5	p = .621
	[k]	7	5	R^2 = .023
1x2	**[p]**	1	10	**p = .001**
	[k]	14	4	**R^2 = .475**
2x1	**[p]**	6	17	**p = .006**
	[k]	17	7	**R^2 = .220**
1x1	**[p]**	2.37	5.00	p = .909
	[k]	4.67	7.22	R^2 = .002

Recognition task following 1x1

Several analyses were computed for the second part of the 1x1 protocol, which consisted in a recognition task. First, we tested whether recognition of a pseudo-word depends on the previous evaluation of its adequacy with a label (i.e. the higher the initial judgement of congruence, the better the recognition). Second, we tested the impact of the 'sound symbolic congruence' (according to our hypotheses) between the target pseudo-word and the label it was presented with in the first part of the protocol, regardless of subjects' judgments. Neither was conclusive. However, the weak number of misrecognitions may be insufficient for analyses, which may be explained by the low difficulty of the task. Only 21% of the pseudo-words that were heard during the first stage were not recognized during the second stage. This percentage of incorrect answers may seem sufficient for the analyses, but it is not given the number of participants and the allotment of these incorrect answers in specific conceptual contrasts. For example, there were only three pseudo-words out of 31 that were presented with the label 'repulsive' and that were not recognized (two [inin] and one [ikik]). This is not enough to obtain satisfying analyses and to reach conclusions about the putative impact of sound symbolism on recognition. As a consequence, we focused on the first part of the 1x1 protocol.

References

Cox, D. R., & Snell, E. J. (1989). *The Analysis of Binary Data (2nd ed.)*. London: Chapman and Hall.

McFadden, D. (1979). Quantitative methods for analysing travel behavior of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), *Behavioural travel modelling* (pp. 279–318). London: Croom Helm.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*(1), 103–120.

Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, *78*(3), 691–692.

Smith, T. J., & Mckenna, C. M. (2013). A Comparison of Logistic Regression Pseudo R^2 Indices. *Multiple Linear Regression Viewpoints*, *39*(2), 17–26.

Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Sonderforschungsbereich*, *386*, Paper 18. Retrieved from http://epub.ub.uni-muenchen.de/