Supplementary Material

The following includes additional information about data collection and analysis.

Data Collection and Coding

The on-line studies were run on Amazon Mechanical Turk using the Crowdflower crowdsourcing service (www.crowdflower.com) with participants recruited from the US, Canada, and Australia. The on-line studies were run using the Qualtrics survey software (www.qualtrics.com[1]).

To be included in the final sample for a given study, participants had to meet the following requirements. (1) Their IP address had not previously appeared in the same or a previous study (either from the series reported here or from related experiments). That is, only the first occurrence of an IP address was used; in the case where the same IP appeared at overlapping times, both sets of responses were discarded. (2) They reported an age of 16 or greater. (3) They answered "yes" to the question: "Is English your first language ("mother tongue")? (4) They answered all questions. (The web program required a response to each question before the participant could progress, so participants without a full response set must have left the task early). Some of the studies included questions asking whether all of the images had displayed properly, and either participants or individual responses were excluded in cases where participants indicated a problem. For Studies 1a and 1b, individual responses were excluded if the reported display time of the "Get Ready" message or either of the two stimuli was out by more than 0.5 seconds. Experiment 2b was a pen-and-paper task; sixty five out of 200 participants who indicated a first language other than English were excluded. For all experiments, the n-values reported in the main text are after these exclusions; the data from excluded participants were not analyzed.

[1] We noted some instability with this widely-used platform. After testing was complete, we ran through the experimental programs and discovered that the appearance of the stimulus was sometimes briefly preceded by the name of the corresponding image file (e.g., "10_1000" for the "Win $10 Win $1000" pair with the small item on the left in Experiment 2a, or "light weight" for the 20 lb weight of Experiment 1a). This error was sporadic, only happening for some runs/trials/web browsers, and did not seem to happen when we first ran the experiments (it may have been a consequence of the image library becoming overloaded, or a change made to the Qualtrics platform). Experiments 1b, 2b, 3b, and 4c were unaffected by this issue because the image files were linked to the Qualtrics software in a different way (and Experiment 2b used a pen-and-paper task). Given that the results of these experiments are identical to those of the potentially-affected experiments, we do not regard the stimulus display problem as having had an important effect on our findings. However, a useful lesson for other researchers using the Qualtrics platform is that image display seems to be much more reliable when the image files are hosted on a local server and linked to via a URL rather than uploaded to the Qualtrics library.

For the on-line studies, the sentence-completion tasks presented the two objects and, beneath them, a sentence such as: "One circle is _____ than the other". Below this were instructions: "In the space below, type the word that you would naturally use to fill the gap in this sentence", followed by a text box into which participants could type their response. In the pen-and-paper version, participants simply wrote their response in the blank space in the middle of the sentence. In Experiments 1a and 1b, where the two objects appeared one after the other, the sentence completion task appeared after the second object and was reworded to be past tense (e.g., "One square was _____ than the other").

For the sentence-completion tasks, responses were coded as "smaller", "larger", or "unclassifiable/irrelevant", with the latter type excluded from analysis on a case-by-case basis. General coding principles included: (1) the response had to include a comparative adjective. That is, if the response is X, one can say "X than..."; non-comparative adjectives (such as "big") were excluded. (2) The adjective must be appropriate for, and clearly refer to, the focal dimension (e.g., a response of "darker" for the area stimuli would be excluded). (3) Modifiers (e.g., "very") were ignored when deciding on the category of a response, as were spelling or grammatical errors and extraneous words (e.g., where the participant typed the whole of the to-be-completed sentence rather than just the missing word). (4) Unusual responses were acceptable provided they could reasonably be taken to refer to the dimension of interest and could be classified as "smaller" or "larger". (5) Affective or value judgments (e.g., "better"), contradictory responses, and ambiguous responses were excluded. Ambiguous responses included ones which were incomprehensible and ones where classification as "smaller" or "larger" was problematic (e.g., "more small").

As noted in the main text, only a small proportion of responses were excluded and inter-rater agreement was excellent. A full copy of the responses from all experiments (and their categorization as "smaller", "larger", and "unclassifiable") is available from the authors.

For the two choice experiments (Experiments 3a and 3b), participants selected which of two words best described the relationship between the items in the stimulus pair. In Experiment 3a the options were the modal "smaller" and "larger" responses for each dimension taken from Experiment 2a (which used identical stimuli). In Experiment 3b the options were: Area: smaller, larger; Height (flags): lower, higher; Length: shorter, longer; Money: less, more; Number: fewer, more; Probability: lower, higher; Time: shorter, longer; Weight: lighter, heavier.

Data Analysis

For the sake of brevity, the main text shows the proportion of "larger" responses collapsed over items. That is, we calculated, for each participant, the proportion of classifiable responses that were coded as "larger" for each experimental condition. (As we note in the main text, each participant only saw one instance of each stimulus pair and randomization/exclusion of unclassifiable responses meant that some participants do not provide data for all conditions.)

Here were present a more complete analysis in which the data for each dimension are shown separately. Figures S1-S3 show the results for each dimension (area, length, height, etc) for Experiments 1a-3b. Each bar shows the proportion of participants in a given condition who responded "larger". These proportions were calculated after excluding the small number of unclassifiable/irrelevant responses; thus, the proportion of "smaller" responses is simply one minus the proportion of larger responses. White bars show the results when the larger item of the pair was presented first (Experiments 1a, 1b) or on the left (Experiments 2a-3b); grey bars show the results when the larger item was presented second or on the right.

The results mirror those in the main text: Across dimensions, there is a robust tendency to favour "larger" comparatives (the HULC effect). In Experiments 1a and 1b, the HULC effect is modulated by the temporal order of the stimuli: "larger" responses were more common when the smaller member of the pair was shown first for all 18 comparisons (10 significant). In Experiments 2a and 2b, the choice of comparative adjective is influenced by the spatial arrangement of the items: The proportion of "larger" responses is always greater when the larger member of the pair is on the left (the white bars are above the grey ones) and this difference is significant for 12/16 comparisons. [Experiments 4a-4c also randomized left-right arrangement and replicated this pattern for 24/24 comparisons (22 significant).] Experiments 3a and 3b suggest that the effect of spatial order is specific to the sentence completion task; only two of the 16 spatial order effects are significant, and overall there is little indication of a systematic effect of spatial order on people's responses in these choice tasks, bolstering the idea that the influence of spatial order in the sentence completion studies reflects a tendency to match word order to object order.

Figure S4 shows the proportion of participants who responded "larger" for each dimension in each condition of Experiments 4a-4c. As before, there is a robust HULC effect across multiple dimensions and stimulus values. More important are the comparisons between magnitude conditions. According to the similarity hypothesis (see main text), there will be more "larger" responses in the big-jump condition (where the items are very different) than in the small-pair and large-pair conditions (where the stimuli are more similar). The top two panels of Figure S5 plot the relevant contrasts (arranged in ascending order in each panel, to clarify the overall pattern). The similarity hypothesis predicts positive differences; there is little indication of this. The big jump vs.

small pair comparison (top panel) has 15/24 contrasts in the predicted direction (2 significant) and 9 in the wrong direction (2 significant); the big jump vs. large pair comparison has just 8/24 contrasts in the predicted direction (1 significant) and 16 in the wrong direction (4 significant). Thus, increasing the similarity of the items does not seem to ameliorate the HULC effect.

By contrast, there is some support for the magnitude hypothesis. The bottom panel of Figure S5 shows that in 18/24 cases participants were more likely to say “larger” when both items were large than when both were small; this difference was significant in 10 cases. Taken together, these studies suggest that the magnitudes of the items, rather than the difference between them, moderate people’s use of comparative adjectives. In all three panels, it is the pair with the highest mean magnitude which is more likely to elicit a “larger” response.

Tables S1-S5 give the absolute number of "smaller" and "larger" responses for each experiment, organized by stimulus dimension and condition (e.g., left-right arrangement). Note that randomization of conditions meant that the total number of responses in each condition was not constant. The tables show the chi-square tests used to establish whether “larger” and “smaller” responses are equiprobable and whether the choice of comparative adjective depends on experimental condition. In all cases, there is one degree of freedom and the critical values are: 3.841 (p <.05), 6.635 (p<.01), and 10.828 (p<.001).

Tables S6 to S8 give the modal responses for each dimension in each experiment. The tables show both the modal "smaller" and "larger" responses, and the overall modal responses.

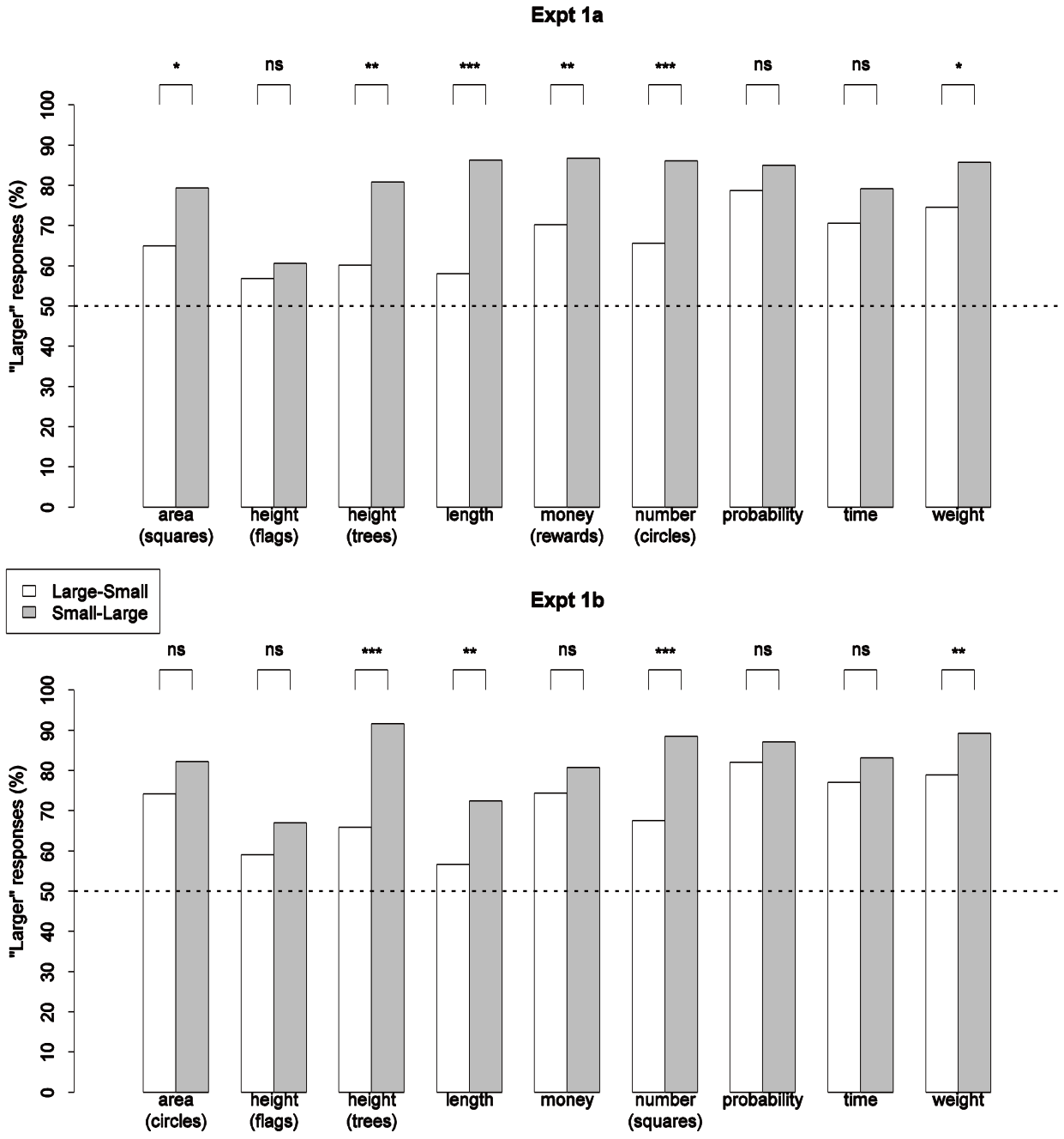

Figure S1. Results from Experiments 1a and 1b. These studies used a sentence-completion task and varied the temporal sequence of the stimuli. The white bars show the proportion of "larger" responses when the larger member of the pair was shown first; gray bars show the proportion of "larger" responses when the larger member of the pair was shown second. Note that all bars are above the 50% line, indicating a robust preference for "larger" responses. Significance markers indicate the results of a chi-squared test for association between response and temporal order: ns = p >.05; * = p <.05; ** = p <.01; *** = p <.001.

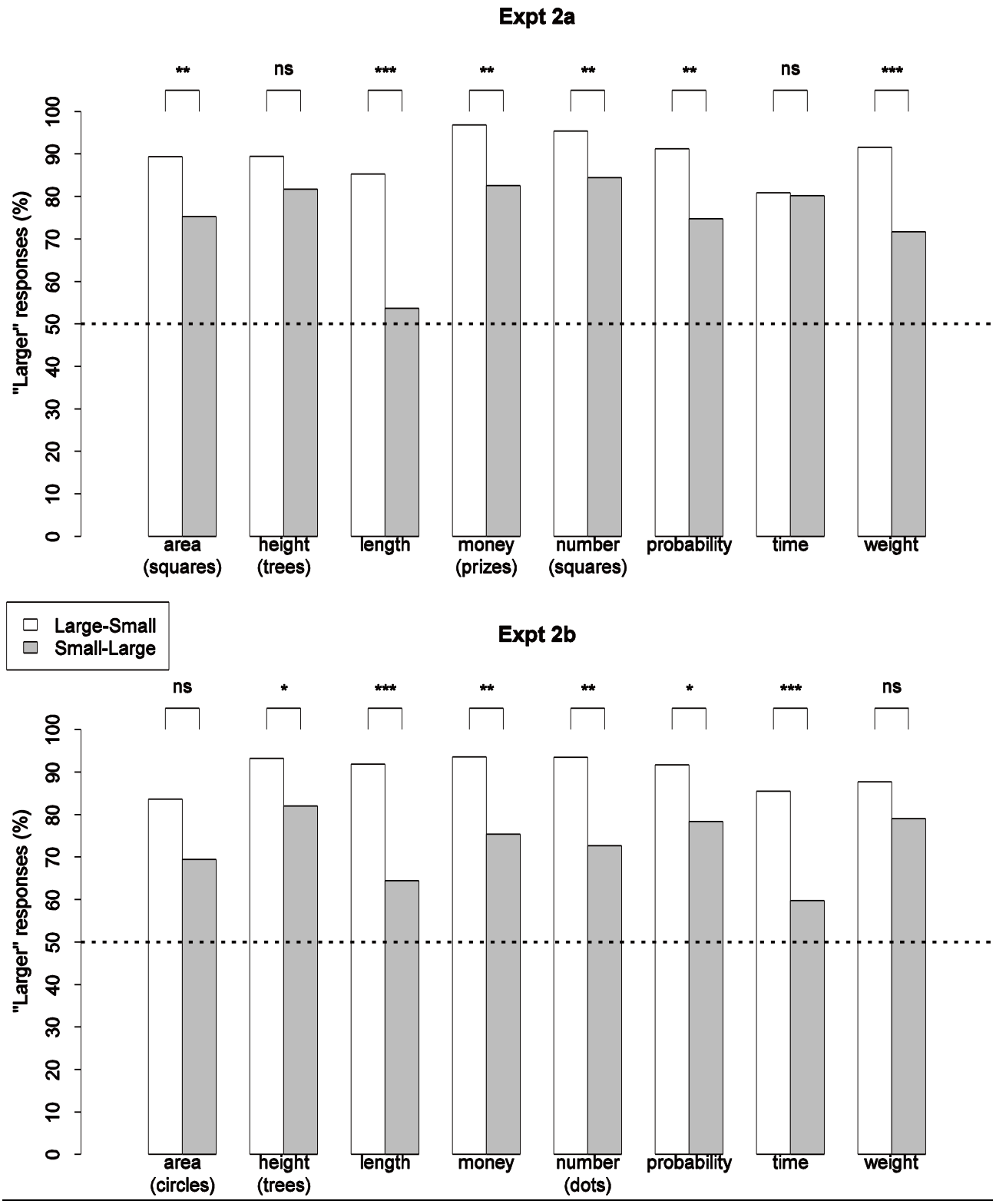

Figure S2. Results from Experiments 2a and 2b. These studies used a sentence completion task and varied the spatial arrangement of the items in each pair. The white bars show the proportion of "larger" responses when the larger member of the pair was on the left; gray bars show the proportion of "larger" responses when the larger member of the pair was on the right. The significance markers indicate the results of a chi-squared test for association between response and spatial arrangement: ns = $p > .05$; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

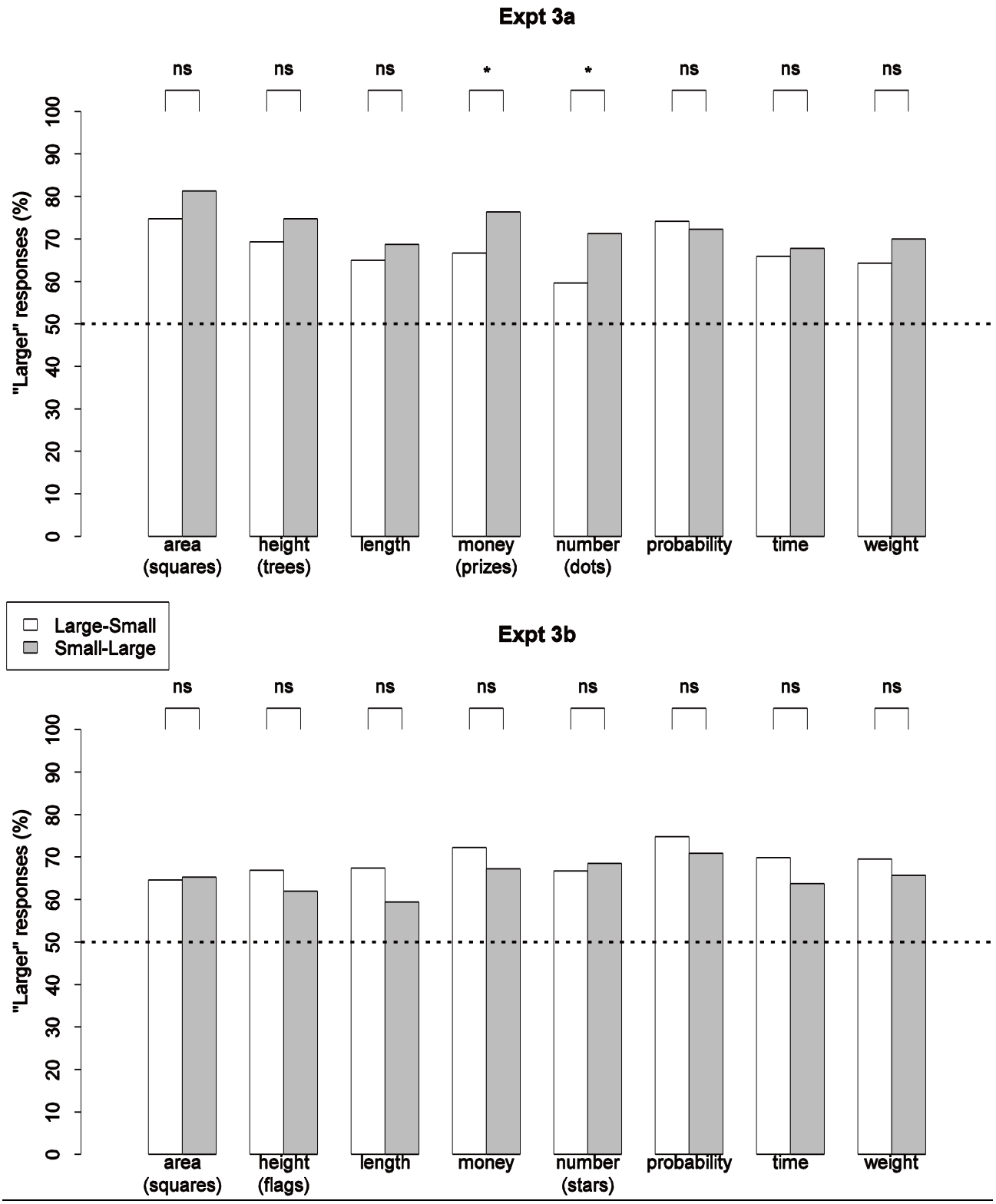

Figure S3. Results from Experiments 3a and 3b. These studies used a two-alternative forced choice task. The white bars show the proportion of "larger" responses when the larger member of the pair was on the left; gray bars show the proportion of "larger" responses when the larger member of the pair was on the right. Significance markers indicate the results of a chi-squared test for association between response and spatial arrangement: ns = p >.05; * = p <.05; ** = p <.01; *** = p <.001.

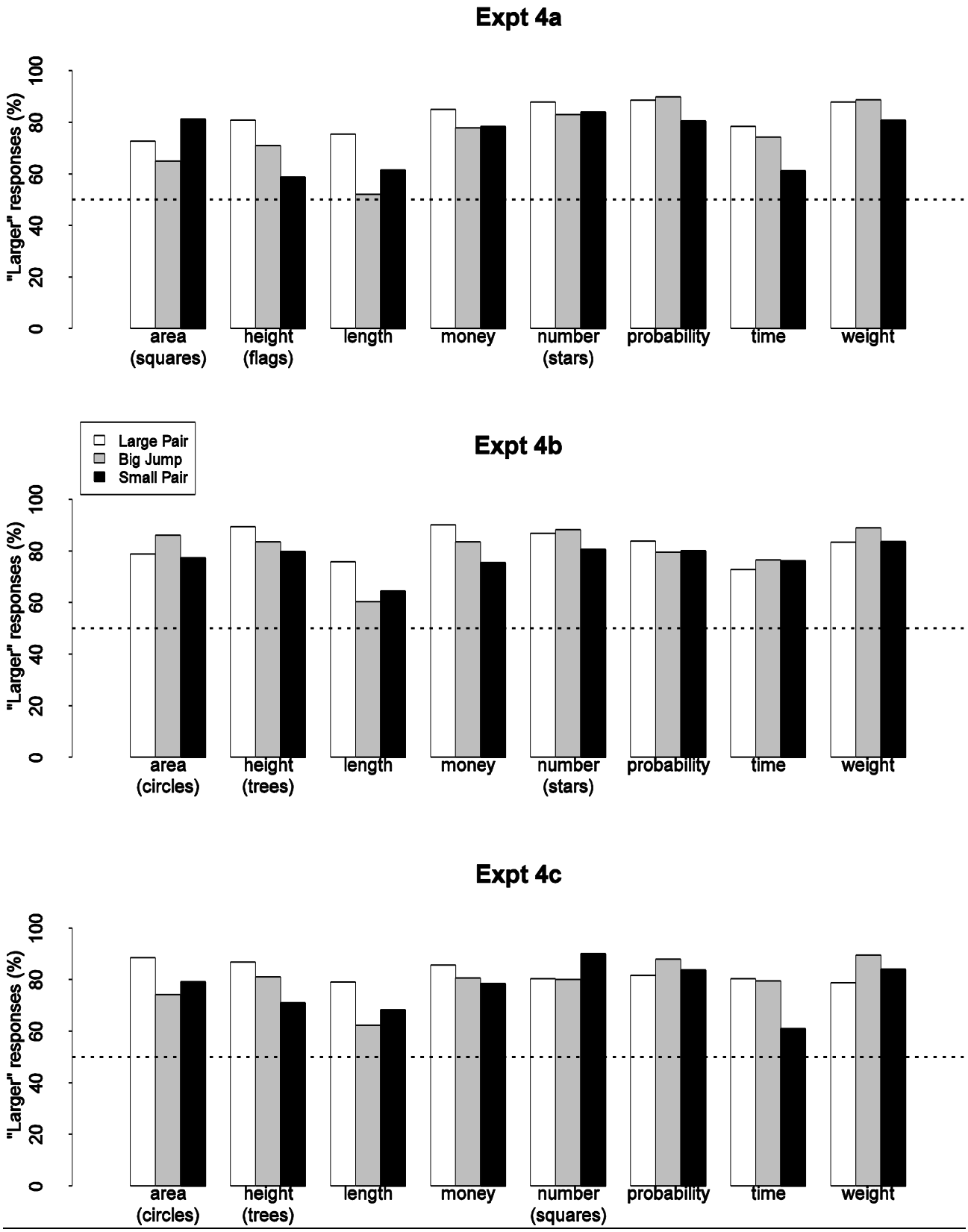

Figure S4. Results from Experiments 4a-4c. White bars show the proportion of "larger" responses when both objects were large (the "large pair"); black bars show the proportion when both objects were small ("small pair"); gray bars show the proportion when there was a big difference between the small and large member of each pair ("big jump").

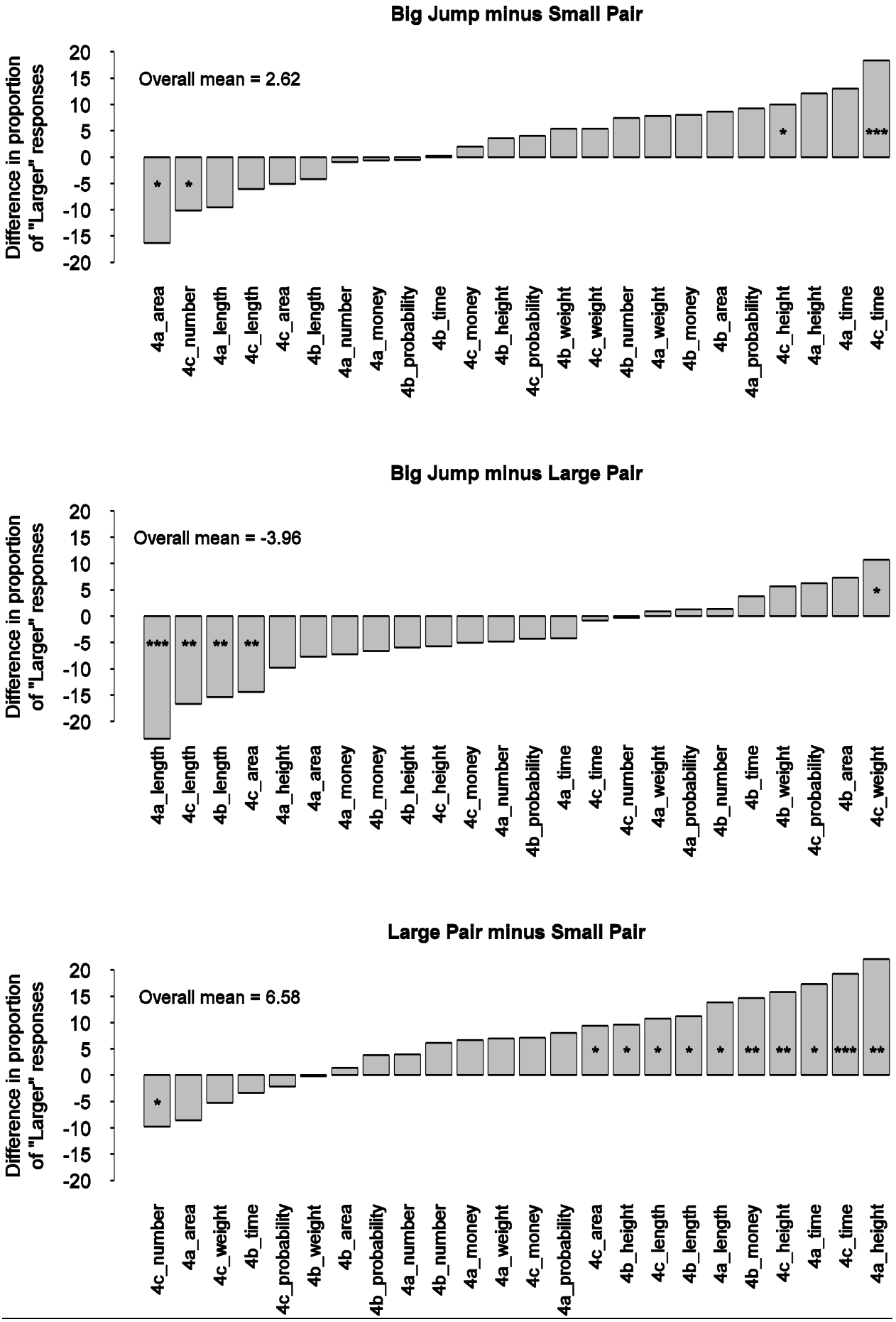

Figure S5. Relevant contrasts for Experiments 4a-4c. The top two panels show the difference in the proportion of "larger" responses between the big-jump condition and the small-pair and large-pair

conditions, respectively. The bottom panel shows the difference between the proportion of “larger” responses for the large-pair and small-pair conditions. In all panels, differences are arranged in order of increasing positivity to clarify the overall pattern. The bar labels indicate the experiment (e.g., 4a) and dimension in question. The significance markers indicate the results of chi-squared tests for association between response type and condition: * = p <.05; ** = p <.01; *** = p <.001.

Table S1. Number of "smaller" and "larger" responses for Experiments 1a and 1b.

		Small-Large			Large-Small			Temporal order
Expt	Dimension	n_S	n_L	χ^2	n_S	n_L	χ^2	χ^2
	Area (squares)	21	81	35.29	35	65	9.00	5.23
	Height (flags)	39	60	4.45	41	54	1.78	0.28
	Height (trees)	22	93	43.83	35	53	3.68	10.52
	Length	13	82	50.12	44	61	2.75	19.49
1a	Money (rewards)	12	79	49.33	33	78	18.24	7.90
	Number (dots)	15	93	56.33	32	61	9.04	11.74
	Probability	15	85	49.00	20	74	31.02	1.29
	Time	22	84	36.26	25	60	14.41	1.91
	Weight	14	84	50.00	28	82	26.51	4.01
	Area (circles)	33	153	77.42	47	135	42.55	3.53
	Height (flags)	57	116	20.12	70	101	5.62	2.36
	Height (trees)	15	166	125.97	61	118	18.15	35.94
	Length	49	129	35.96	81	106	3.34	9.91
1b	Money	35	147	68.92	46	134	43.02	2.08
	Number (squares)	20	154	103.20	60	125	22.84	22.70
	Probability	21	142	89.82	30	137	68.56	1.63
	Time	29	143	75.56	42	141	53.56	2.06
	Weight	20	166	114.60	38	142	60.09	7.36

Note: Small-Large and Large-Small refer to the temporal order of the to-be-compared objects. n_S and n_L are the number of "smaller" and "larger" responses, respectively; the corresponding chi-square values test whether these two types of response occurred equally often. The final column gives the chi-square value for a test of association between response type ("smaller" vs. "larger") and temporal order.

Table S2. Number of "smaller" and "larger" responses for Experiments 2a and 2b.

		Small-Large			Large-Small			Spatial order
Expt	Dimension	n_S	n_L	χ^2	n_S	n_L	χ^2	χ^2
2a	Area (squares)	27	82	27.75	10	84	58.26	6.76
	Height (trees)	19	85	41.88	10	85	59.21	2.39
	Length	50	58	0.59	14	81	47.25	23.32
	Money (prizes)	15	71	36.47	3	92	83.38	10.28
	Number (dots)	14	76	42.71	5	104	89.92	6.87
	Probability	24	71	23.25	8	83	61.81	8.85
	Time (delays)	20	81	36.84	18	76	35.79	0.01
	Weight	30	76	19.96	8	87	65.69	12.92
2b	Area (circles)	19	43	9.29	12	61	32.89	3.83
	Height (trees)	11	50	24.93	5	68	54.37	3.95
	Length	26	47	6.04	5	56	42.64	14.05
	Money	18	55	18.75	4	58	47.03	8.15
	Number (dots)	20	53	14.92	4	57	46.05	9.82
	Probability	13	47	19.27	6	66	50.00	4.72
	Time	29	43	2.72	9	53	31.23	10.88
	Weight	13	49	20.90	9	64	41.44	1.83

Note: Small-Large and Large-Small refer to the left-right arrangement of the to-be-compared objects. n_S and n_L are the number of "smaller" and "larger" responses, respectively; the corresponding chi-square values test whether these two types of response occurred equally often. The final column gives the chi-square value for a test of association between response type ("smaller" vs. "larger") and spatial arrangement.

Table S3. Number of "smaller" and "larger" responses for Experiments 3a and 3b.

		Small-Large			Large-Small			Spatial order
Expt	Dimension	n_S	n_L	χ^2	n_S	n_L	χ^2	χ^2
3a	Area (squares)	33	143	68.75	47	139	45.51	2.23
	Height (trees)	47	139	45.51	54	122	26.27	1.32
	Length	57	125	25.41	63	117	16.20	0.55
	Money (prizes)	43	139	50.64	60	120	20.00	4.19
	Number (dots)	52	129	32.76	73	108	6.77	5.39
	Probability	51	133	36.54	46	132	41.55	0.16
	Time (delays)	57	120	22.42	63	122	18.82	0.14
	Weight	54	126	28.80	65	117	14.86	1.34
3b	Area (squares)	66	124	17.71	67	122	16.01	0.02
	Height (flags)	72	117	10.71	63	127	21.56	1.01
	Length	78	114	6.75	61	126	22.59	2.61
	Money	63	129	22.69	52	135	36.84	1.12
	Number (stars)	59	128	25.46	64	128	21.33	0.14
	Probability	54	131	32.05	49	145	47.51	0.74
	Time	69	121	14.23	57	132	29.76	1.62
	Weight	66	126	18.75	57	130	28.50	0.66

Note: Columns as are in Table S2.

Table S4. Number of "smaller" and "larger" responses in Experiments 4a-4c.

Expt	Dimension	S1S2		S2S1		L1L2		L2L1		S1L2		L2S1	
		n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L
4a	Area (squares)	12	27	6	51	19	37	10	40	23	26	11	37
	Height (flags)	22	25	18	32	8	34	7	29	16	27	9	34
	Length	27	27	15	40	17	32	5	35	30	15	18	37
	Money	18	43	6	44	10	44	5	41	10	30	8	33
	Number (stars)	7	37	7	36	10	50	3	44	14	43	3	40
	Probability	13	32	4	38	10	45	2	47	5	30	3	40
	Time	18	23	10	21	17	47	7	40	18	28	5	38
	Weight	12	35	7	45	10	33	2	53	8	37	3	49
4b	Area (circles)	19	38	7	51	20	49	7	51	17	51	1	60
	Height (trees)	15	41	8	50	7	52	7	66	17	49	4	57
	Length	33	32	11	48	27	39	7	67	29	29	15	38
	Money	19	36	10	53	7	49	5	60	16	43	6	68
	Number (stars)	21	41	1	51	9	50	8	62	12	64	3	48
	Probability	18	46	5	46	12	44	8	59	20	38	3	51
	Time	18	36	8	47	25	41	8	47	24	46	4	45
	Weight	15	48	7	64	16	35	3	60	11	58	3	55
4c	Area (circles)	23	50	8	68	8	55	8	69	27	48	11	61
	Height (trees)	22	50	18	48	13	61	6	64	18	59	11	65
	Length	28	36	16	59	19	53	11	60	44	33	14	63
	Money	25	48	7	69	14	56	7	69	21	46	6	66
	Number (squares)	8	76	6	51	20	55	8	59	22	62	8	58
	Probability	15	64	9	60	16	52	9	59	11	52	5	64
	Time	29	40	22	40	18	47	7	55	28	57	3	63
	Weight	18	56	7	76	21	56	8	52	12	57	3	71

Note: Column headings indicate stimulus pair and left-right arrangement (e.g., S1S2 means S1 was on the left and S2 was on the right). Thus, S1S2 and S2S1 refer to the two left-right arrangements of the "small pair"; L1L2 and L2L1 refer to the two arrangements of the "large pair"; and S1L2 and L2S1 are the two left-right orientations of the "big jump" stimuli.

Table S5. Chi-square values for Experiments 4a-4c.

Expt	Dimension	SL	LS	Spatial Order	Small Pair vs Big Jump	Large Pair vs Big Jump	Small Pair vs Large Pair
4a	Area (squares)	9.00	65.81	15.24	6.51	1.40	2.09
	Height (flags)	12.12	28.84	2.21	2.95	2.15	9.72
	Length	0.00	36.51	19.32	1.91	10.94	4.27
	Money	40.26	71.54	5.25	0.01	1.57	1.53
	Number (stars)	60.88	86.08	5.14	0.03	0.98	0.62
	Probability	46.23	100.42	11.15	2.76	0.07	2.36
	Time	13.41	49.00	9.63	3.13	0.49	6.40
	Weight	41.67	114.62	12.84	2.33	0.04	1.79
4b	Area (circles)	34.66	122.08	24.87	3.08	2.36	0.06
	Height (trees)	58.61	123.52	9.63	0.53	1.95	4.38
	Length	0.64	77.42	36.79	0.43	6.81	3.96
	Money	43.51	126.73	13.44	2.49	2.40	9.03
	Number (stars)	64.82	128.33	15.29	2.59	0.11	1.68
	Probability	34.18	113.95	20.18	0.01	0.72	0.56
	Time	16.51	89.06	23.80	0.00	0.44	0.35
	Weight	53.56	143.52	19.60	1.60	1.62	0.00
4c	Area (circles)	42.77	129.96	16.64	1.05	9.76	4.66
	Height (trees)	61.39	95.11	3.55	4.04	1.82	10.62
	Length	4.51	89.15	30.57	1.16	9.90	4.15
	Money	38.57	151.14	27.81	0.19	1.29	2.52
	Number (squares)	84.15	112.19	6.23	5.74	0.00	5.36
	Probability	75.60	124.27	6.16	0.96	2.03	0.23
	Time	21.74	83.56	15.95	11.52	0.03	11.49
	Weight	63.29	150.97	18.21	1.91	6.02	1.34

Note: Column SL gives the chi-square values testing whether "smaller" and "larger" responses were equally likely when the large item was on the right (small-large layout). In all cases bar one there is a significant preference for use of "larger" comparatives. Column LS gives the same values for the large-small layout. Larger comparatives are significantly more likely in almost every case. The Spatial Order column gives the chi-square test for an association between response and left-right arrangement. The last three columns give the chi-square tests for association used to see whether the proportion of "smaller" and "larger" responses depended on the magnitudes of the presented objects.

Table S6. Modal responses in Experiments 1a and 1b.

Expt		"Smaller"	"Larger"	Overall	*n*
1a	Area (squares)	smaller	bigger	bigger	202
	Height (flags)	lower	higher	higher	194
	Height (trees)	shorter	taller	taller	203
	Length	shorter	longer	longer	200
	Money (rewards)	less	larger	larger	202
	Number (dots)	fewer	more	more	201
	Probability	smaller	greater	greater	194
	Time	shorter	longer	longer	191
	Weight	lighter	heavier	heavier	208
1b	Area (circles)	smaller	bigger	bigger	368
	Height (flags)	lower	higher	higher	344
	Height (trees)	shorter	taller	taller	360
	Length	shorter	longer	longer	365
	Money	less	more	more	362
	Number (squares)	fewer	more	more	359
	Probability	smaller	greater	greater	330
	Time	shorter	longer	longer	355
	Weight	lighter	heavier	heavier	366

Note: The "Smaller" and "Larger" columns give the most common responses among those classified as "smaller" and "larger"; the Overall column gives the most common response ignoring category. The *n* values are the sample sizes for each dimension after removing unclassifiable responses.

Table S7. Modal responses in Experiments 2a and 2b.

Expt		"Smaller"	"Larger"	Overall	*n*
2a	Area (squares)	smaller	larger	larger	203
	Height (trees)	shorter	taller	taller	199
	Length	shorter	longer	longer	203
	Money (prizes)	smaller	bigger	bigger	181
	Number (dots)	fewer	more	more	199
	Probability	less	greater	greater	186
	Time (delays)	shorter	longer	longer	195
	Weight	lighter	heavier	heavier	201
2b	Area (circles)	smaller	bigger	bigger	135
	Height (trees)	shorter	taller	taller	134
	Length	shorter	longer	longer	134
	Money	less	greater	greater	135
	Number (dots)	less	more	more	134
	Probability	less	greater	greater	132
	Time	shorter	longer	longer	134
	Weight	less	heavier	heavier	135

Table S8. Modal responses in experiments 4a-4c.

Expt		"Smaller"	"Larger"	Overall	n
4a	Area (squares)	smaller	bigger	bigger	299
	Height (flags)	lower	higher	higher	261
	Length	shorter	longer	longer	298
	Money	less	more	more	292
	Number (stars)	less	more	more	294
	Probability	less	greater	greater	269
	Time	shorter	longer	longer	272
	Weight	less	heavier	heavier	294
4b	Area (circles)	smaller	bigger	bigger	371
	Height (trees)	shorter	taller	taller	373
	Length	shorter	longer	longer	375
	Money	less	larger	larger	372
	Number (stars)	less	more	more	370
	Probability	less	greater	greater	350
	Time	shorter	longer	longer	349
	Weight	less	heavier	heavier	375
4c	Area (circles)	smaller	larger	larger	436
	Height (trees)	shorter	taller	taller	435
	Length	shorter	longer	longer	436
	Money	less	more	more	434
	Number (squares)	less	more	more	433
	Probability	less	greater	greater	416
	Time	shorter	longer	longer	409
	Weight	lighter	heavier	heavier	437