Appendix A: Corpus details

Attention check questions

In addition to the 300 sentences below, which formed the basis of the analyses in this paper, each experiment incorporated some attention-check questions. These were intended to be used to identify participants who were not paying attention to the task or did not understand it. As such, they were two pairs of sentences with a clear answer, as shown below (the first sentence of the pair is the better one).

- Mike read things quickly. / Read things, Mike did quickly.
- Eva was killed by John. / Eva was killed from John.

In the INITIAL / WITHIN PARTICIPANTS two-session experiment format, in TARGET PAIRS and RANDOM PAIRS choice tasks, both pairs of sentences were presented in both halves of the experiment. In the LIKERT and ME rating tasks, one sentence from each pair was presented in the first half and the other sentence from each was presented in the second half. A participant was considered to have failed the attention check if they chose the incorrect sentence over the correct one in a choice task or ranked an incorrect sentence more highly than a correct one in a rating task.

The BETWEEN PARTICIPANTS single-session format was identical except that no sentences appeared in the second half (since there was no second half). Participant exclusions were determined before any other analyses and the attention check questions were not included in the main analysis for any measure.

0.1 Instruction quiz

Participants in all conditions had to pass a two-question instruction quiz testing their understanding of acceptability as defined in the instructions (reported in the main *Method* section. These may have influenced their interpretation of the task. The questions asked were:

Which of these best describes "sentence acceptability", the thing you'll be asked to make judgments about?
- How likely the sentence is to offend someone.
- Whether the sentence expresses positive or negative sentiment.
- How "well formed" the sentence is, whether it sounds natural.
- How likely the sentence is to be true.

How can you tell us about the acceptability of these sentences?

- You'll be asked to give sentences an acceptability rating.
- You'll be asked to compare two sentences and say which one is more acceptable
- You'll be asked to describe what is wrong with some badly worded sentences.
- Both of options 1 and 2, but not 3.

Expected answers were options 3 and 4 respectively.

Main sentences

The sentences used here were a randomly selected subset of those collected by Sprouse et al. (2013). Although drawing randomly from the instances presented for each phenomenon, we did sample uniformly across the 150 sources present in the full collection when drawing the set of 150 pairs of sentences used here. Sprouse et al. (2013) identify repeated phenomena among this set, one example appearing in our subset is the pair of sentences *I seem eating sushi / I like eating sushi* and the pair *I tend taking vacations / I like taking vacations*, which Sprouse et al. (2013) consider equivalent. In addition to direct repeats, we expect further theoretical connections to be present among related but non-identical phenomena, but for the purposes of the analyses presented above the critical property of this set of sentences is just that they cover a range of acceptability values representative of linguistics research. As noted in the main text, multiple items testing a single phenomenon are necessary to make reliable claims regarding its status: occasional repeats notwithstanding, in general we chose to sacrifice the ability to make claims about the status of individual phenomena in order to increase the diversity of the contrasts considered, since the latter is more directly relevant to the test-retest measure of reliability.

Pair Id	Proposed as acceptable	Proposed less acceptable
1	Brittany attempted to touch the porcupine.	Brittany touched plenty of.
2	It seems to him that Kim solved the problem.	He seems to that Kim solved the problem.
3	Scott intended to run for class president.	Scott intended to have run for class president.
4	Debra convinced Elliot that he would wash the dishes.	Debra convinced Elliot to have washed the dishes.

5	Celia convinced Brad that he will have eaten dinner by the time she gets home.	Celia convinced Brad that he would have eaten dinner by the time she gets home.
6	Charlene believed Shawn to be helpful.	Charlene believed Shawn to write the paper.
7	Thomas read in the paper that the stock market crashed.	Thomas read in the paper the stock market crashed.
8	It is important for one to sleep regularly.	It is important one to sleep regularly.
9	The expectation that Lauren will graduate is reasonable.	The expectation Lauren will graduate is reasonable.
10	How likely to sell the house is Diane?	How likely to be a stock market crash is there?
11	Carla appears to have graduated high school.	Carla was hoped to graduate high school.
12	"Stop bullying me!" shouted the overweight child fearfully.	"Stop bullying me!" shouted fearfully the overweight child.
13	Jason drove his car, and Tara rode her bike.	Jason drove his car, and Tara her bike rode.
14	Annie was insulted.	Was insulted Annie.
15	The money was stolen.	The money was stolen the money.
16	Beth hitchhiked to Los Angeles and Robert drove to San Diego.	Beth hitchhiked Los Angeles and Robert drove to San Diego.
17	Richard may have been hiding, but Blake may have been doing so too.	Richard may have been hiding, but Blake may have done so too.
18	They all have eaten and they have all done so quickly.	They all have eaten and they have done all so quickly.
19	Arty is sick, and Mark is too.	Arty is sick, and Mark does so too.
20	John tried to win.	John tried himself to win.
21	Dale found Brooke after frightening himself.	Dale found Brooke after frightening herself.
22	Olivia told Gregory when to exonerate himself.	Olivia told Gregory when to exonerate herself.
23	Last month there was a plan to promote me.	Last month there was a plan to promote oneself.
24	Edward created a website in order for us to promote ourselves.	Edward created a website in order for us to promote himself.
25	Melissa's pledge to Dan to take care of herself.	Melissa's pledge to Dan to take care of himself.

26	Eric shouted to Maria to believe in herself.	Eric shouted to Maria to believe in himself.
27	Jack asked Sally to be allowed to take care of himself.	Jack asked Sally to be allowed to take care of herself.
28	She called Victor and yelled at him.	She called Victor and yelled at.
29	There has been an announcement made in the newspaper.	There has been made an announcement in the newspaper.
30	There has been a man considered violent.	There has been considered a man violent.
31	She wrote her boyfriend a poem.	She wrote a poem her boyfriend.
32	We believed Ben with all our hearts to be innocent.	We believed with all our hearts Ben to be innocent.
33	We consider there to be three problems.	We consider there three problems.
34	I told Mandy who would win the election.	I told Mandy who the hell would win the election.
35	During no storm should people leave their computers on.	What during no storm should people leave on?
36	I don't think that I will take any musicians to the office.	I don't think that any musicians, I will take to the office.
37	Who the hell asked who out?	Who asked who the hell out?
38	Clare didn't promise Rita a red cent.	Clare didn't promise every employee a red cent.
39	I visited a city yesterday near the city that John did.	I visited a city near the city yesterday that John did.
40	I told you when we met that Bill will come to the party.	I told you that Bill when we met will come to the party.
41	I suggested when we dated that Laura should cut her hair.	I suggested that Laura when we dated should cut her hair.
42	Melanie prefers for everyone to get a raise that you do.	Melanie prefers for everyone you do to get a raise.
43	It rained torrentially.	It torrentially rained.
44	Mike believes Phil to be a genius.	Mike believes to be a genius.
45	There are likely to be students in the library.	There are likely students to be in the library.
46	There might seem to be fossils several miles underground.	There might fossils seem to be several miles underground.
47	The flowers were planted to attract hummingbirds.	The flowers plant seasonally to attract hummingbirds.

48	The bureaucrat was bribed deliberately.	The bureaucrat bribes deliberately.
49	The book was written truthfully.	The was written book truthfully.
50	Picking which suit to wear in the morning makes Helen late to work.	Picking who to wear that new suit in the morning makes late to work.
51	Someone better leave town.	Anyone better leave town.
52	John wondered what he bought.	What did John wonder what he bought?
53	Which coworker did George compliment before insulting?	Which coworker did George yawn before insulting?
54	It appears that tonight Marjorie is staying over.	Tonight appears that Marjorie is staying over.
55	Who did I see hug Natalie?	Who was seen hug Natalie?
56	There are leaves burnt.	There are leaves green.
57	The doctors are almost all wealthy.	The doctors almost all are wealthy.
58	What do you complain that the neighbor turns on at night?	What do you notice if the neighbor turns on at night?
59	It appears that a certain player will leave the team, but which player is still a secret.	It appears that a certain player will leave the team, but which player it does is still a secret.
60	The professor said that the score of one the assignments is going to be posted online by the end of the day, but I don't remember which.	The professor said that the score of one the assignments is going to be posted online by the end of the day, but I don't remember which he did.
61	Melissa said she read about one of Shakespeare's plays, but I don't know which play.	Melissa said she read about one of Shakespeare's plays, but I don't know which play she did.
62	Richard and Christine hate each other.	Each other hate Richard and Christine.
63	Sabrina gave her elderly mother a beautiful new coffee mug at breakfast time.	Sabrina gave at breakfast time her elderly mother a beautiful new coffee mug.
64	The poor were overlooked by the politicians and the bankers.	The poor were overlooked and the bankers by the politicians.
65	The government gives no help to any poor people.	The government gives any help to no poor people.
66	Julian learned to drive, and Rita learned to compose music.	Julian learned to drive, and Rita did to compose music.

67	The instructor put more solutions than the TA did on the board.	The instructor put more solutions on the board than the TA did on the handout.
68	John intended to give the children something nice to eat, and give the children a generous handful of candy he did.	John intended to give the children something nice to eat, and give the children he did a generous handful of candy.
69	The table was being set by the waiter.	The table was being heavy.
70	We supporters of democrats are just as worried about the economy as you supporters of republicans.	We supporters of democrats are just as worried about the economy as you of republicans.
71	He was the assistant.	He was assistant.
72	The brother and sister that were playing all the time had to be sent to bed.	Brother and sister that were playing all the time had to be sent to bed.
73	The hammer with the black handle and the screwdriver with the square tip are in your toolbox.	Hammer with the black handle and screwdriver with the square tip are in your toolbox.
74	The book is long and the essay is short.	Book is long and essay is short.
75	This is a shelf.	This is shelf.
76	The lobby of the movie theater with the fantastic sound system was empty.	The movie theater with the fantastic sound system's lobby was empty.
77	There seems to be a new deal in the works.	There desires to be a new deal in the works.
78	One interpreter tried to be assigned to every visiting diplomat.	One interpreter each tried to be assigned to every visiting diplomat.
79	Nobody has gone anywhere.	Anywhere has nobody gone.
80	The virtuoso only rarely practices any pieces.	The virtuoso practices any pieces only rarely.
81	It seemed at that time that Robert had confessed.	It seemed at that time Robert had confessed.
82	What they children believe is that they will get some candy.	What they children believe is they will get some candy.
83	They denied and we suspected that Sean would buy the car.	They denied and we suspected Sean would buy the car.

84	Matt believed that Ben read a book and Lilly that Ben watched a movie.	Matt believed Ben read a book and Lilly Ben watched a movie.
85	That Addison bit the boy, Jena didn't believe.	Addison bit the boy, Jena didn't believe.
86	At that time, what did they believe that Peter fixed?	What did they believe at that time that Peter fixed?
87	This is the child who I think will walk your dog.	This is the child who I think that will walk your dog.
88	Eric wondered how Lisa learned to dance a certain dance, but it's not clear what dance.	Eric wondered how Lisa learned to dance, but it's not clear what dance.
89	They consider a teacher of Chris geeky.	Who do they consider a teacher of geeky?
90	Cassie is more talented than intelligent.	Cassie is more than intelligent talented.
91	They anticipate that everybody will contact Fred that you do.	They anticipate that everybody you do will contact Fred.
92	Fran searched the web for Danny.	Fran searched Danny the web.
93	The eerie sound frightened Seth.	The eerie sound frightened Seth at the ghost.
94	I find it annoying that usually this bus is late.	I find it annoying for usually this bus to be late.
95	A dog bigger than my corgi started snarling.	A bigger dog than my corgi started snarling.
96	Jessica shouted at a girl as nervous as her daughter.	Jessica shouted at as nervous a girl as her daughter.
97	The heat turned the meat rotten.	The heat turned the meat rotted.
98	It seems a woman is in the yard.	It seems a woman to be in the yard.
99	There is likely to appear a man.	There is likely a man to appear.
100	It is unimaginable for Mary to arrive on time.	It is unimaginable Mary to arrive on time.
101	What did you contribute to whom?	To whom did you contribute what?
102	Some frogs and a fish are in the pond.	Some frogs and a fish is in the pond.
103	The more you give, the happier you will be.	That much the more you give, the happier you will be.
104	Brian asked who sent what.	Brain asked what who sent.

105	He met Bush, about whom he wrote several stories.	He met about whom he wrote several stories.
106	He would have been fired.	Him would have been fired.
107	There tend to be storms at this time of year.	There like to be storms at this time of year.
108	I like eating sushi.	I seem eating sushi.
109	I like taking vacations.	I tend taking vacations.
110	The tennis players watched the ball bounce a crazy bounce off the line.	My parents appeared an unexpected appearance last night.
111	How much money is there in your account?	How much money there is in your account?
112	Into which room walked three men?	Into which room did walk three men?
113	Onto which floor did the wine glass fall?	Onto which floor fell the wine glass?
114	Mr. Ted said we should write five pages each night, but the actual doing of it turned out to be impossible.	Mr. Ted said we should write five pages each night, but the actual doing of so turned out to be impossible.
115	From the Atlantic to the Pacific there are people who oppose immigration reform.	With the generals about a new approach the president talked.
116	It will take five to seven days for your cat to feel better.	It will take from five seven days for your cat to feel better.
117	The newlyweds travelled a few days and then settled down at a motel.	A few days, the newlyweds travelled, and then settled down at a motel.
118	Larry cooked her husband the meal	Who did Larry cook the meal?
119	There seems to be a box on your doorstep.	There seems a box to be on your doorstep.
120	Dale was suspended by the principal.	Dale was suspended from the principal.
121	The freezer was defrosted to remove the ice build-up.	The freezer defrosted to remove the ice build-up.
122	The chamber was flooded intentionally.	The chamber flooded intentionally.
123	John broke a cup, and Mary did so too.	John broke a cup, and Mary did so with a saucer.
124	Linda graduated high school before her brother.	Linda graduated high school her brother.

125	Leonard talks about the things Sally likes.	Leonard talks what Sally likes.
126	Eric began a romantic relationship with an employee who he later found himself superior to.	Eric began a romantic relationship with an employee superior to who later found himself.
127	Who brought what?	What did who bring?
128	He envied me my success after the promotion.	He envied me after the promotion my success.
129	Maggie is difficult to compliment without embarrassing.	Maggie is unlikely to be complimented without anyone embarrassing.
130	That Mary was going out with Luke bothered you.	Who did that Mary was going out with bother you?
131	Angela wonders which story about Elaine is online.	Who does Angela wonder which story about is online?
132	The teacher denied George his extra credit.	The teacher denied his extra credit to George.
133	In that room anyone who stays long enough is given horrible headaches.	In that room is given anyone who stays long enough horrible headaches.
134	The prom queen was picked some flowers before the award ceremony.	Before the award ceremony was picked the prom queen some flowers.
135	The child that the clown gave the creeps to at the party is still upset.	The child that the clown gave the creeps at the party to is still upset.
136	The children are certain to have all been picked up.	All the children are certain to have all been picked up.
137	Shannon stole food for her family to enjoy.	Shannon stole for her family to enjoy.
138	The question was answered feeling nervous.	The question was answered nervous.
139	I informed Stan that I wanted to wash the car together.	I informed Stan that I must wash the car together.
140	The gardener and handyman painted the fence together.	The gardener painted the fence together.
141	There have arisen in these negotiations all of the issues the lawyers warned us about.	There have all arisen in these negotiations the issues the lawyers warned us about.
142	Caroline likes cupcakes and Lisa cookies.	Caroline likes cupcakes because Lisa cookies.

143	Kelsey speaks Japanese more fluently than Jason does.	Kelsey speaks Japanese more fluently than Jason English.
144	The girls made sandwiches, and I believe that the boys did too.	The girls made sandwiches, and I believe that the boys hot dogs.
145	The bully threatened students triumphantly.	Threaten students, the bully did triumphantly.
146	Sarah expected to receive a good grade.	Expect to receive, Sarah did a good grade.
147	Who did the executives appoint a friend of to the board?	Who did the executives appoint a friend of chairman of the board?
148	When yesterday Marnie started to present that report, I thought we were in for a lot of surprises.	When that report Marnie started to present yesterday, I thought we were in for a lot of surprises.
149	Bill knows that such books John only reads at home.	Bill asked if such books John only reads at home.
150	If George comes, the party will be a disaster.	If George probably comes, the party will be a disaster.

Proposed as acceptable	Proposed less acceptable	Target pairs Initial/within	Target pairs Between	Likert Initial/within	Likert Between
Celia convinced Brad that he will have eaten dinner by the time she gets home.	Celia convinced Brad that he would have eaten dinner by the time she gets home.	∼,×	∼	∼,∼	∼
How likely to sell the house is Diane?	How likely to be a stock market crash is there?	×,∼	×	∼,∼	∼
Richard may have been hiding, but Blake may have been doing so too.	Richard may have been hiding, but Blake may have done so too.	×,∼	×	∼,∼	∼
We believed Ben with all our hearts to be innocent.	We believed with all our hearts Ben to be innocent.	×,×	×	✓,✓	∼
Melanie prefers for everyone to get a raise that you do.	Melanie prefers for everyone you do to get a raise.	×,×	∼	∼,∼	∼
There are leaves burnt.	There are leaves green.	×,×	∼	✓,✓	✓
What do you complain that the neighbor turns on at night?	What do you notice if the neighbor turns on at night?	×,×	∼	∼,✓	∼
The instructor put more solutions than the TA did on the board.	The instructor put more solutions on the board than the TA did on the handout.	×,×	×	×,∼	∼
It seemed at that time that Robert had confessed.	It seemed at that time Robert had confessed.	×,×	∼	✓,✓	∼
What they children believe is that they will get some candy.	What they children believe is they will get some candy.	×,×	×	∼,∼	∼
They denied and we suspected that Sean would buy the car.	They denied and we suspected Sean would buy the car.	∼,×	×	✓,∼	∼

That Addison bit the boy, Jena didn't believe.	Addison bit the boy, Jena didn't believe.	×,~	×	~,~	~
Eric wondered how Lisa learned to dance a certain dance, but it's not clear what dance.	Eric wondered how Lisa learned to dance, but it's not clear what dance.	×,×	×	~,~	~
They anticipate that everybody will contact Fred that you do.	They anticipate that everybody you do will contact Fred.	×,×	×	~,~	~
A dog bigger than my corgi started snarling.	A bigger dog than my corgi started snarling.	~,×	✓	~,~	~
There is likely to appear a man.	There is likely a man to appear.	×,×	~	~,~	~
What did you contribute to whom?	To whom did you contribute what?	×,~	×	~,~	~
Who did the executives appoint a friend of to the board?	Who did the executives appoint a friend of chairman of the board?	~,~	×	~,~	~
Bill knows that such books John only reads at home.	Bill asked if such books John only reads at home.	~,~	×	✓,✓	✓

This table lists only those pairs of sentences for which there is disagreement in the decisions made by different measures. Expert judgments from *Linguistic Inquiry* are taken as the reference standard, ✓indicates that the formal measure endorses the expert-preferred item, × indicates the formal measure endorses the less preferred item, and ~ indicates the formal measure failed to reject the null hypothesis of no difference. Two relevant formal measures are given, TARGET PAIRS for its distinctively high power and LIKERT to represent the consensus of the other formal measures. Since this data set does not in general contain multiple items targeting the same phenomena, little can be said about any possible relation between particular sentence structures and the behavior of any one measure. What is important from a test-retest reliability viewpoint is that there are different ways in which disagreements can arise. Some appear to be simple sampling error, which seems to be the most likely explanation for the single instance of disagreement with expert judgment in the *Celia convinced Brad...* sentence pair (pair #1). Others

appear to be instances where the high power of TARGET PAIRS allows it to pick up a counterintuitive effect too small to be detected with the other measures, such as *The instructor...* (pair #8). Finally, there do appear to be rare instances where presenting controlled contrasts directs people's attention to features of a construction that are not salient when the sentences involved are considered in isolation or contrasted with unrelated partners, such as *There are leaves...* (pair #6) or *... Robert had confessed* (pair #9).

Appendix B: Details of the THURSTONE model

Choice task responses

Choice task responses were modeled using a hierarchical variation of the traditional Thurstonian model, fit using JAGS (Plummer 2003) as outlined in Figure 1. Participants responded to a series of sentence pairs. For each pair, three response options were available: one for each of the sentences indicating that sentence is more acceptable, plus one to indicate no difference. Given a collection of responses, the model infers an acceptability strength for each item under the assumptions that the language community is homogeneous, and that acceptability is transitive.

The acceptability strength scores are inferred using the following decision process mapping acceptability strength to choice responses: When presented with test items a and b on trial j, participant i samples an acceptability strength for each item. By definition, these samples come from normal distributions with means centered at the group's consensual acceptability score for each item, α_a and α_b. Each participant has their own degree of variability around the shared mean, σ_i. The participant then takes the difference between the perceived acceptability scores on this trial, L_{ij}, and compares it to their criterion level C_i. The criterion determines how each participant maps their perception of the acceptability difference onto the response options. If the absolute difference is greater than the criterion value, the participant indicates the preference indicated by the sign of the difference, otherwise they indicate no preference. Although different in some respects, this mapping process follows similar motivating logic to a scheme previously described by Bader & Häussler (2010) for mapping between graded sentence acceptability responses and binary ones.
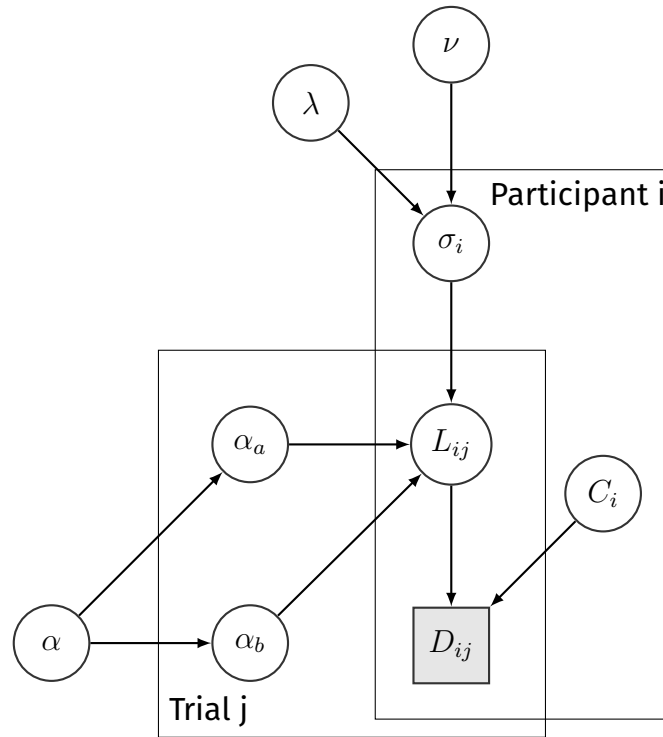
Figure 1: Model diagram showing the implementation of the THURSTONE model. Variable definitions are given in Table 3 .

Variable	Meaning	Distributed as
D_{ij}	Observed decision	$\begin{cases} 1 & \text{if } L_{ij} > C, \\ -1 & \text{if } L_{ij} < -C, \\ 0 & \text{otherwise} \end{cases}$
L_{ij}	Decision variable for trial j	$N(\alpha_a - \alpha_b, \sigma_i)$
C	Participant decision criterion	$U(1, 50)$
σ	Participant variability	$Weibull(\lambda, \nu)$
λ	Variability hyperparameter	$\lvert N(2, 5) \rvert$
ν	Variability hyperparameter	$\lvert N(0.02, 0.04) \rvert$
α	Item acceptability	$N(0, 20)$

Table 3: Parameter descriptions.

References

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330.

Plummer, Martyn. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, vol. 124, 125.

Sprouse, Jon, Carson Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua* 134. 219–248.