

Category	Standard	Prior-trained	<i>p</i> -value
agreement_determiner_noun-across_1_adjective <i>look at this purple thing .</i> <i>*look at this purple things .</i>	0.85 ± 0.00	0.85 ± 0.00	0.00
agreement_determiner_noun-between_neighbors <i>this color must be white .</i> <i>*this colors must be white .</i>	0.88 ± 0.01	0.88 ± 0.01	0.56
agreement_subject_verb-across_prepositional_phrase <i>the lie on the foot is flat .</i> <i>*the lies on the foot is flat .</i>	0.62 ± 0.01	0.62 ± 0.01	0.85
agreement_subject_verb-across_relative_clause <i>the book that i like is poor .</i> <i>*the books that i like is poor .</i>	0.65 ± 0.01	0.67 ± 0.01	0.00
agreement_subject_verb-in_question_with_aux <i>where does the horse go ?</i> <i>*where does the horses go ?</i>	0.76 ± 0.01	0.76 ± 0.01	0.06
agreement_subject_verb-in_simple_question <i>where is the way ?</i> <i>*where is the ways ?</i>	0.82 ± 0.00	0.80 ± 0.01	0.00
anaphor_agreement-pronoun_gender <i>will Mark want himself ?</i> <i>*will Mark want herself ?</i>	0.90 ± 0.01	0.88 ± 0.01	0.01
argument_structure-dropped_argument <i>give me the poor boat .</i> <i>*the poor boat gives me .</i>	0.89 ± 0.00	0.91 ± 0.00	0.00

Table 1: Zorro results (1/3). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
argument_structure-swapped_arguments <i>the wolf made her ball they .</i> <i>*they made the wolf her ball .</i>	0.98 \pm 0.00	1.00 \pm 0.00	0.00
argument_structure-transitive <i>Philip thinks .</i> <i>*Philip affected .</i>	0.65 \pm 0.00	0.68 \pm 0.00	0.00
binding-principle-a <i>Ben thinks about himself calling this fuel .</i> <i>*Ben thinks about himself called this fuel .</i>	0.82 \pm 0.01	0.77 \pm 0.01	0.00
case-subjective_pronoun <i>i brought the wolf my hill .</i> <i>*the wolf brought i my hill .</i>	0.98 \pm 0.01	1.00 \pm 0.00	0.00
ellipsis-n_bar <i>Mark fixed one worn canal and Roger fixed more .</i> <i>*Mark fixed one canal and Roger fixed more worn .</i>	0.49 \pm 0.01	0.50 \pm 0.01	0.57
filler-gap-wh_question_object <i>Laura married the dinner that the wolf could close .</i> <i>*Laura married what the dinner could close the wolf .</i>	0.96 \pm 0.00	0.97 \pm 0.00	0.02
filler-gap-wh_question_subject <i>Laura ended the finger that can make boats .</i> <i>*Laura ended who the finger can make boats .</i>	0.98 \pm 0.00	0.98 \pm 0.00	0.75
irregular-verb <i>Michael chose the good one some time ago .</i> <i>*Michael chosen the good one some time ago .</i>	0.82 \pm 0.01	0.84 \pm 0.01	0.00

Table 2: Zorro results (2/3). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
island-effects-adjunct_island <i>who should William have without watching the baby ?</i> <i>*who should William have the baby without watching ?</i>	0.71 ± 0.01	0.66 ± 0.02	0.00
island-effects-coordinate_structure_constraint <i>who must Philip and the dinosaur turn ?</i> <i>*who must Philip turn and the dinosaur ?</i>	0.71 ± 0.02	0.77 ± 0.03	0.00
local_attractor-in_question_with_aux <i>is the whale getting the person ?</i> <i>*is the whale gets the person ?</i>	0.97 ± 0.00	0.96 ± 0.00	0.00
npi_licensing-matrix_question <i>does her boat ever play with the growth ?</i> <i>*her boat does ever play with the growth ?</i>	0.66 ± 0.02	0.68 ± 0.02	0.08
npi_licensing-only_npi_licensor <i>only Mark ever finds some suit .</i> <i>*even Mark ever finds some suit .</i>	0.62 ± 0.04	0.76 ± 0.04	0.00
quantifiers-existential_there <i>there are many books about soft birds .</i> <i>*there are most books about soft birds .</i>	0.91 ± 0.01	0.91 ± 0.01	0.16
quantifiers-superlative <i>no pig could stand on top of more than six days .</i> <i>*no pig could stand on top of at least six days .</i>	0.44 ± 0.04	0.49 ± 0.03	0.05
overall	0.79 ± 0.00	0.80 ± 0.00	0.00

Table 3: Zorro results (3/3). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
adjunct_island <i>What was Bill keeping after cleaning this book?</i> * <i>What was Bill keeping this book after cleaning?</i>	0.77 \pm 0.01	0.80 \pm 0.01	0.00
anaphor_gender_agreement <i>Patricia had found herself.</i> * <i>Patricia had found itself.</i>	0.74 \pm 0.01	0.71 \pm 0.01	0.00
anaphor_number_agreement <i>Lisa shouldn't remember herself.</i> * <i>Lisa shouldn't remember themselves.</i>	0.88 \pm 0.01	0.86 \pm 0.01	0.00
animate_subject_passive <i>Every cafe is found by the best friends.</i> * <i>Every cafe is found by the boxes.</i>	0.42 \pm 0.01	0.39 \pm 0.01	0.00
animate_subject_trans <i>Adam should take a light.</i> * <i>The gate should take a light.</i>	0.76 \pm 0.00	0.79 \pm 0.00	0.00
causative <i>Walter hasn't wrinkled a skirt.</i> * <i>Walter hasn't fallen a skirt.</i>	0.69 \pm 0.00	0.72 \pm 0.00	0.00
complex_NP_island <i>Who didn't every French teacher who isn't working with some brothers of the doctor hear?</i> * <i>Who didn't some brothers of the doctor hear every French teacher who isn't working with?</i>	0.45 \pm 0.01	0.44 \pm 0.01	0.01
coordinate_structure_constraint_complex_left_branch <i>What bases were a lot of brothers arguing about and those students thinking about?</i> * <i>What were a lot of brothers arguing about bases and those students thinking about?</i>	0.55 \pm 0.01	0.60 \pm 0.01	0.00
coordinate_structure_constraint_object_extraction <i>What would Mark and Grace shake?</i> * <i>What could Mark shake and Grace?</i>	0.60 \pm 0.01	0.59 \pm 0.01	0.02
coordinate_structure_constraint_subject_extraction <i>Who and Alice had those ladders confused?</i> * <i>Who had those ladders confused and Alice?</i>	0.21 \pm 0.02	0.14 \pm 0.01	0.00

Table 4: BLiMP_{CH} results (1/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	p -value
determiner_noun_agreement_1 <i>Martha didn't notice that waiter.</i> * <i>Martha didn't notice that waiters.</i>	0.78 ± 0.00	0.81 ± 0.00	0.00
determiner_noun_agreement_2 <i>Martha passed that river.</i> * <i>Martha passed those river.</i>	0.92 ± 0.00	0.94 ± 0.00	0.00
determiner_noun_agreement_irregular_1 <i>Marcus wouldn't hate those teeth.</i> * <i>Marcus wouldn't hate those tooth.</i>	0.84 ± 0.01	0.86 ± 0.00	0.00
determiner_noun_agreement_irregular_2 <i>Partners hadn't attacked those men.</i> * <i>Partners hadn't attacked that men.</i>	0.96 ± 0.00	0.98 ± 0.00	0.00
determiner_noun_agreement_with_adj_2 <i>Kristin breaks this purple glass.</i> * <i>Kristin breaks these purple glass.</i>	0.86 ± 0.00	0.85 ± 0.00	0.01
determiner_noun_agreement_with_adjective_1 <i>Denise wouldn't visit these tiny hills.</i> * <i>Denise wouldn't visit these tiny hill.</i>	0.78 ± 0.00	0.77 ± 0.00	0.02
determiner_noun_agreement_with_adj_irregular_1 <i>Every guest cleaned these thin feet.</i> * <i>Every guest cleaned these thin foot.</i>	0.89 ± 0.00	0.88 ± 0.00	0.00
determiner_noun_agreement_with_adj_irregular_2 <i>Some piano teachers keep that smart goose.</i> * <i>Some piano teachers keep those smart goose.</i>	0.90 ± 0.01	0.90 ± 0.00	0.01
distractor_agreement_relational_noun <i>The father of many patients has looked to attack the guest.</i> * <i>The father of many patients have looked to attack the guest.</i>	0.27 ± 0.01	0.24 ± 0.01	0.00
distractor_agreement_relative_clause <i>That niece that remembered many ladies wasn't buying those sweaters.</i> * <i>That niece that remembered many ladies weren't buying those sweaters.</i>	0.30 ± 0.01	0.26 ± 0.00	0.00

Table 5: BLiMP_{CH} results (2/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
drop_argument <i>Charles visited.</i> * <i>Charles holds.</i>	0.54 ± 0.00	0.54 ± 0.00	0.60
ellipsis.n_bar_1 <i>Students throw no strange dishes and a lot of women throw some.</i> * <i>Students throw no dishes and a lot of women throw some strange.</i>	0.70 ± 0.01	0.72 ± 0.01	0.00
ellipsis.n_bar_2 <i>The doctor grows four tomatoes and Leslie grows five blue tomatoes.</i> * <i>The doctor grows four rough tomatoes and Leslie grows five blue.</i>	0.83 ± 0.00	0.87 ± 0.00	0.00
existential.there.object.raising <i>Rachel could find there to be an actor eating some pasta.</i> * <i>Rachel could beg there to be an actor eating some pasta.</i>	0.67 ± 0.01	0.66 ± 0.01	0.12
existential.there.quantifiers_1 <i>There were some stairs disappearing.</i> * <i>There were all stairs disappearing.</i>	0.94 ± 0.01	0.92 ± 0.00	0.00
existential.there.quantifiers_2 <i>All public parks weren't there disturbing Thomas.</i> * <i>There weren't all public parks disturbing Thomas.</i>	0.77 ± 0.02	0.84 ± 0.01	0.00
existential.there.subject.raising <i>There weren't bound to be many hats drying.</i> * <i>There weren't unable to be many hats drying.</i>	0.58 ± 0.01	0.55 ± 0.01	0.00
expletive.it.object.raising <i>Ann's friends will find it to be not so nice that Suzanne was hurting Carmen.</i> * <i>Ann's friends force it to be not so nice that Suzanne was hurting Carmen.</i>	0.61 ± 0.01	0.61 ± 0.01	0.61
inchoative <i>That door would shut.</i> * <i>That door would love.</i>	0.60 ± 0.00	0.60 ± 0.00	0.31
intransitive <i>This patient's teacher came.</i> * <i>This patient's teacher heard.</i>	0.51 ± 0.00	0.53 ± 0.00	0.00

Table 6: BLiMP_{CH} results (3/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
irregular_past_participle_adjectives <i>The forgotten partners aren't sick.</i> * <i>The forgot partners aren't sick.</i>	0.69 ± 0.03	0.70 ± 0.02	0.23
irregular_past_participle_verbs <i>Lori wrote.</i> * <i>Lori written.</i>	0.79 ± 0.02	0.78 ± 0.01	0.49
irregular_plural_subject_verb_agreement_1 <i>The man was bothering every patient.</i> * <i>The man are bothering every patient.</i>	0.81 ± 0.01	0.81 ± 0.00	0.71
irregular_plural_subject_verb_agreement_2 <i>The men skate around these hills.</i> * <i>The man skate around these hills.</i>	0.76 ± 0.01	0.75 ± 0.01	0.02
left_branch_island_echo_question <i>Dana has noticed what waiter?</i> * <i>What has Dana noticed waiter?</i>	0.79 ± 0.01	0.76 ± 0.01	0.00
left_branch_island_simple_question <i>Which father was this report disgusting?</i> * <i>Which was this report disgusting father?</i>	0.74 ± 0.01	0.75 ± 0.01	0.13
matrix_question_npi_licensor_present <i>Has Samantha ever stood up?</i> * <i>Samantha has ever stood up.</i>	0.26 ± 0.01	0.28 ± 0.01	0.00
npi_present_1 <i>Even most dresses certainly twisted.</i> * <i>Even most dresses ever twisted.</i>	0.18 ± 0.01	0.22 ± 0.01	0.00
npi_present_2 <i>Guy often washes.</i> * <i>Guy ever washes.</i>	0.22 ± 0.01	0.23 ± 0.01	0.05
only_npi_licensor_present <i>Only Thomas has ever pressed this French teacher to sound like Sandra.</i> * <i>Even Thomas has ever pressed this French teacher to sound like Sandra.</i>	0.60 ± 0.03	0.74 ± 0.04	0.00

Table 7: BLiMP_{CH} results (4/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
only_npi_scope <i>Only every child who floors do help did ever fall asleep.</i> <i>*Every child who only floors do help did ever fall asleep.</i>	0.85 ± 0.01	0.84 ± 0.01	0.35
passive_1 <i>The mother of William wasn't helped by a father of Jennifer.</i> <i>*The mother of William wasn't gone fishing by a father of Jennifer.</i>	0.69 ± 0.01	0.71 ± 0.01	0.01
passive_2 <i>Sabrina isn't bothered.</i> <i>*Sabrina isn't cried.</i>	0.63 ± 0.01	0.63 ± 0.01	0.48
principle_A_case_1 <i>Stephen hadn't expected that he broke those plates.</i> <i>*Stephen hadn't expected that himself broke those plates.</i>	1.00 ± 0.00	1.00 ± 0.00	0.65
principle_A_case_2 <i>Daniel forgot about himself hiding every closet.</i> <i>*Daniel forgot about himself hides every closet.</i>	0.80 ± 0.01	0.77 ± 0.01	0.00
principle_A_c_command <i>The associates that escaped from Keith don't forgive themselves.</i> <i>*The associates that escaped from Keith don't forgive himself.</i>	0.64 ± 0.01	0.60 ± 0.01	0.00
principle_A_domain_1 <i>Derek wouldn't agree that Amelia teaches him.</i> <i>*Derek wouldn't agree that Amelia teaches himself.</i>	0.98 ± 0.00	0.99 ± 0.00	0.00
principle_A_domain_2 <i>Linda won't know Vincent hasn't met himself.</i> <i>*Linda won't know Vincent hasn't met herself.</i>	0.68 ± 0.01	0.66 ± 0.01	0.00
principle_A_domain_3 <i>Lisa writes Laurie's grandfather hides himself.</i> <i>*Laurie's grandfather writes Lisa hides himself.</i>	0.62 ± 0.00	0.61 ± 0.00	0.09
principle_A_reconstruction <i>It's himself that George understood.</i> <i>*It's himself that understood George.</i>	0.48 ± 0.01	0.40 ± 0.01	0.00

Table 8: BLiMP_{CH} results (5/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
regular_plural_subject_verb_agreement_1 <i>Some doctors are bringing a cart.</i> <i>*Some doctors was bringing a cart.</i>	0.64 \pm 0.01	0.65 \pm 0.01	0.06
regular_plural_subject_verb_agreement_2 <i>The sister isn't going to the rivers.</i> <i>*The sisters isn't going to the rivers.</i>	0.64 \pm 0.01	0.64 \pm 0.01	0.58
sentential_negation_npi_licensor_present <i>Sandra can not ever meet a lot of guests.</i> <i>*Sandra can really ever meet a lot of guests.</i>	0.96 \pm 0.00	0.98 \pm 0.00	0.00
sentential_negation_npi_scope <i>Holly's fathers that are scaring all doctors would not ever speak up.</i> <i>*Holly's fathers that are not scaring all doctors would ever speak up.</i>	0.64 \pm 0.02	0.73 \pm 0.02	0.00
sentential_subject_island <i>Who were most adults' bothering Guy distracting.</i> <i>*Who were most adults' bothering distracting Guy.</i>	0.43 \pm 0.01	0.40 \pm 0.01	0.00
superlative_quantifiers_1 <i>No painting looks like more than seven rugs.</i> <i>*No painting looks like at least seven rugs.</i>	0.41 \pm 0.03	0.44 \pm 0.02	0.22
superlative_quantifiers_2 <i>That dress has confused at most six guests.</i> <i>*No dress has confused at most six guests.</i>	0.81 \pm 0.02	0.87 \pm 0.01	0.00
tough_vs_raising_1 <i>A husband wasn't fun to work with.</i> <i>*A husband wasn't certain to work with.</i>	0.62 \pm 0.01	0.62 \pm 0.01	0.52
tough_vs_raising_2 <i>Renee wasn't about to drive this plane.</i> <i>*Renee wasn't hard work to drive this plane.</i>	0.49 \pm 0.01	0.49 \pm 0.01	0.90
transitive <i>Connie cares for Christopher.</i> <i>*Connie escaped Christopher.</i>	0.72 \pm 0.00	0.75 \pm 0.00	0.00

Table 9: BLiMP_{CH} results (6/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
wh_island	0.62 ± 0.01	0.59 ± 0.01	0.00
<i>What hasn't Melanie found out she goes to?</i>			
* <i>What hasn't Melanie found out who goes to?</i>			
wh_questions_object_gap	0.73 ± 0.01	0.73 ± 0.01	0.82
<i>Many piano teachers question these best friends that every partner noticed.</i>			
* <i>Many piano teachers question who every partner noticed these best friends.</i>			
wh_questions_object_gap_long_distance	0.74 ± 0.01	0.73 ± 0.01	0.45
<i>Some associate couldn't investigate mothers that that mouth that opens hurts.</i>			
* <i>Some associate couldn't investigate who that mouth that opens hurts mothers.</i>			
wh_questions_subject_gap	0.90 ± 0.01	0.90 ± 0.01	0.05
<i>Those students see the books that hurt Alice.</i>			
* <i>Those students see who the books hurt Alice.</i>			
wh_questions_subject_gap_long_distance	0.97 ± 0.00	0.99 ± 0.00	0.00
<i>Connie noticed those cars that that doctor couldn't clean that worry that company.</i>			
* <i>Connie noticed what those cars that that doctor couldn't clean worry that company.</i>			
wh_vs_that_no_gap	0.94 ± 0.00	0.94 ± 0.00	0.04
<i>Christina hasn't found out that a lot of piano teachers breed many rabbits.</i>			
* <i>Christina hasn't found out what a lot of piano teachers breed many rabbits.</i>			
wh_vs_that_no_gap_long_distance	0.95 ± 0.00	0.97 ± 0.00	0.00
<i>Gregory might know that all men that have whispered shock a grandfather.</i>			
* <i>Gregory might know who all men that have whispered shock a grandfather.</i>			
wh_vs_that_with_gap	0.35 ± 0.01	0.34 ± 0.01	0.41
<i>Joseph hasn't remembered what these teachers like.</i>			
* <i>Joseph hasn't remembered that these teachers like.</i>			
wh_vs_that_with_gap_long_distance	0.12 ± 0.01	0.10 ± 0.01	0.00
<i>Valerie wasn't finding out what the paintings that hadn't looked like the drawings look like.</i>			
* <i>Valerie wasn't finding out that the paintings that hadn't looked like the drawings look like.</i>			
overall	0.67 ± 0.00	0.67 ± 0.00	0.00

Table 10: BLiMP_{CH} results (7/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
adjunct_island <i>what does that person keep while dropping those hats ?</i> <i>*what does that person keep those hats while dropping ?</i>	0.88 ± 0.01	0.88 ± 0.01	0.80
adv_order <i>your friend found your teacher quietly .</i> <i>*your friend found quietly your teacher .</i>	0.89 ± 0.01	0.91 ± 0.01	0.02
anaphor_gender_agreement <i>your brother is n't thinking about himself .</i> <i>*your brother is n't thinking about herself .</i>	0.96 ± 0.00	0.94 ± 0.01	0.00
anaphor_number_agreement <i>these children heard themselves .</i> <i>*these children heard himself .</i>	0.86 ± 0.01	0.85 ± 0.02	0.14
center_embedding_double_1 <i>my boss your cousins these teachers care for scare sings .</i> <i>*my boss your cousins these teachers care for sings .</i>	0.30 ± 0.00	0.30 ± 0.00	0.58
center_embedding_double_2 <i>this adult my cousins your friends praise listen to talks .</i> <i>*this adult my cousins your friends praise listen to cry talks .</i>	0.96 ± 0.01	0.99 ± 0.00	0.00
center_embedding_single_1 <i>my friend my friends scare reads .</i> <i>*my friend my friends reads .</i>	0.33 ± 0.01	0.33 ± 0.01	0.54
center_embedding_single_2 <i>my cousins the dancer understands cough .</i> <i>*my cousins the dancer understands sings cough .</i>	0.97 ± 0.00	0.99 ± 0.00	0.00
complex_np_island <i>who did those doctors who can n't care for the guys hear ?</i> <i>*who did the guys hear those doctors who can n't care for ?</i>	0.35 ± 0.01	0.33 ± 0.01	0.03
coordinate_structure_constrain_object_extraction <i>who could Julie and Rachel answer ?</i> <i>*who could Julie answer and Rachel ?</i>	0.61 ± 0.02	0.63 ± 0.03	0.38

Table 11: SCaMP_{plausible} results (1/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
determiner_noun_agreement_1 <i>the women worked with this child .</i> <i>*the women worked with this children .</i>	0.93 \pm 0.00	0.94 \pm 0.00	0.00
determiner_noun_agreement_2 <i>the boy had bothered these men .</i> <i>*the boy had bothered this men .</i>	0.98 \pm 0.00	0.98 \pm 0.00	0.00
determiner_noun_agreement_adj_1 <i>the children were n't breaking this old car .</i> <i>*the children were n't breaking this old cars .</i>	0.92 \pm 0.00	0.92 \pm 0.00	0.18
determiner_noun_agreement_adj_2 <i>the teacher did n't talk about these boring chairs .</i> <i>*the teacher did n't talk about this boring chairs .</i>	0.92 \pm 0.00	0.92 \pm 0.00	0.82
ellipsis <i>the people have made five old forks , and these men have made two .</i> <i>*the people have made five forks , and these men have made two old .</i>	0.71 \pm 0.01	0.75 \pm 0.01	0.00
green_ideas <i>confused doctors stretch quietly .</i> <i>*stretch quietly doctors confused .</i>	0.83 \pm 0.01	0.83 \pm 0.01	0.51
intransitive <i>those doctors smile .</i> <i>*those doctors like .</i>	0.70 \pm 0.01	0.70 \pm 0.01	0.94
left_branch_island_simple_question <i>which car have the doctors noticed ?</i> <i>*which have the doctors noticed car ?</i>	0.76 \pm 0.01	0.75 \pm 0.01	0.07
matrix_question_npi_licensor_present <i>can the teacher ever talk ?</i> <i>*the teacher can ever talk .</i>	0.58 \pm 0.02	0.71 \pm 0.02	0.00
npi_present <i>this man will always praise the dancer .</i> <i>*this man will ever praise the dancer .</i>	0.44 \pm 0.01	0.43 \pm 0.01	0.26

Table 12: SCaMP_{plausible} results (2/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
only_npi_licensor_present <i>only that dancer could ever work with that doctor .</i> <i>*even that dancer could ever work with that doctor .</i>	0.68 ± 0.05	0.79 ± 0.05	0.00
only_npi_scope <i>only the mirror that the driver had dropped should ever roll away .</i> <i>*the mirror that only the driver had dropped should ever roll away .</i>	0.97 ± 0.01	0.97 ± 0.01	0.95
pp_order <i>my teacher lost my teacher on the couch .</i> <i>*my teacher lost on the couch my teacher .</i>	0.85 ± 0.01	0.84 ± 0.01	0.29
principle_A_c_command <i>your grandmother that hears these doctors could hate herself .</i> <i>*your grandmother that hears these doctors could hate themselves .</i>	0.51 ± 0.01	0.49 ± 0.01	0.00
principle_A_domain_1 <i>your daughter should n't explain that my friends notice her .</i> <i>*your daughter should n't explain that my friends notice herself .</i>	1.00 ± 0.00	1.00 ± 0.00	0.00
principle_A_domain_2 <i>Sara says that that king had remembered himself .</i> <i>*Sara says that that king had remembered herself .</i>	0.60 ± 0.01	0.60 ± 0.01	0.87
principle_A_domain_3 <i>this woman had n't said that this man had n't thought about himself .</i> <i>*this man had n't said that this woman had n't thought about himself .</i>	0.50 ± 0.00	0.50 ± 0.00	0.21
recursion_intensifier_adj_long <i>the dentist was very very very very very very very very very very nice .</i> <i>*the dentist sings very very very very very very very very very very nice .</i>	0.75 ± 0.02	0.77 ± 0.02	0.10
recursion_intensifier_adj_medium <i>the adult is very very very very very very hungry .</i> <i>*the adult reads very very very very very very hungry .</i>	0.83 ± 0.01	0.84 ± 0.01	0.45
recursion_intensifier_adj_short <i>the customer is happy .</i> <i>*the customer reads happy .</i>	0.95 ± 0.00	0.95 ± 0.00	0.57

Table 13: SCaMP_{plausible} results (3/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	p-value
recursion_intensifier_adv_long <i>the driver exercises really really really really really really really perfectly .</i> <i>*the driver is really really really really really really really perfectly .</i>	0.84 ± 0.02	0.91 ± 0.02	0.00
recursion_intensifier_adv_medium <i>the woman sings really really really quietly .</i> <i>*the woman is really really really quietly .</i>	0.93 ± 0.01	0.96 ± 0.01	0.00
recursion_intensifier_adv_short <i>the man sings really really slowly .</i> <i>*the man was really really slowly .</i>	0.91 ± 0.01	0.92 ± 0.01	0.37
recursion_poss_ditransitive_long <i>the dentist bought my father 's mother 's sister 's friend 's teacher 's cousin 's mother 's daughter the chair .</i> <i>*the dentist dropped my father 's mother 's sister 's friend 's teacher 's cousin 's mother 's daughter the chair .</i>	0.55 ± 0.04	0.63 ± 0.04	0.01
recursion_poss_ditransitive_medium <i>the child sent my brother 's doctor 's dentist 's friend the shirt .</i> <i>*the child cleaned my brother 's doctor 's dentist 's friend the shirt .</i>	0.57 ± 0.03	0.66 ± 0.03	0.00
recursion_poss_ditransitive_short <i>your dentist bought my cousin the car .</i> <i>*your dentist dropped my cousin the car .</i>	0.76 ± 0.02	0.85 ± 0.02	0.00
recursion_poss_transitive_long <i>Leslie knows that my son 's wife 's son 's dad 's mother 's sister 's daughter 's mom 's doctor 's mother is bored .</i> <i>*Leslie talks to my son 's wife 's son 's dad 's mother 's sister 's daughter 's mom 's doctor 's mother is bored .</i>	0.79 ± 0.04	0.87 ± 0.04	0.00
recursion_poss_transitive_medium <i>Julia knows that your mom 's father 's daughter 's father 's cousin 's grandmother 's cousin is clever .</i> <i>*Julia works with your mom 's father 's daughter 's father 's cousin 's grandmother 's cousin is clever .</i>	0.94 ± 0.01	0.97 ± 0.01	0.01
recursion_poss_transitive_short <i>Denise thinks that your father is tired .</i> <i>*Denise works with your father is tired .</i>	1.00 ± 0.00	1.00 ± 0.00	0.01
recursion_pp_is_long <i>the hat on the chair in the park on the couch by the bag on the street on the table in the school on the ground was awesome .</i> <i>*the hat is on the chair in the park on the couch by the bag on the street on the table in the school on the ground was awesome .</i>	0.87 ± 0.03	0.91 ± 0.04	0.10

Table 14: SCaMP_{plausible} results (4/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
recursion_pp_is_medium <i>the book in the kitchen on the bed on the chair on the stool is red .</i> <i>*the book was in the kitchen on the bed on the chair on the stool is red .</i>	0.89 ± 0.03	0.93 ± 0.03	0.03
recursion_pp_is_short <i>the chair is old .</i> <i>*the chair was is old .</i>	0.96 ± 0.01	0.98 ± 0.01	0.00
recursion_pp_verb_long <i>the glass sitting by the box in the school by the shoe in the room by the bicycle on the ground in the restaurant is dry .</i> <i>*the glass sits by the box in the school by the shoe in the room by the bicycle on the ground in the restaurant is dry .</i>	0.45 ± 0.06	0.52 ± 0.09	0.19
recursion_pp_verb_medium <i>the shoe sitting on the table on the table by the plate in the kitchen in the hospital in the library is purple .</i> <i>*the shoe sits on the table on the table by the plate in the kitchen in the hospital in the library is purple .</i>	0.59 ± 0.06	0.66 ± 0.08	0.17
recursion_pp_verb_short <i>the button sitting is red .</i> <i>*the button sits is red .</i>	0.83 ± 0.03	0.88 ± 0.03	0.03
sentential_negation_npi_licensor_present <i>this woman can n't ever answer these teachers .</i> <i>*this woman can possibly ever answer these teachers .</i>	1.00 ± 0.00	1.00 ± 0.00	0.52
sentential_negation_npi_scope <i>the doctor who has cleaned the toys had n't ever waved .</i> <i>*the doctor who has n't cleaned the toys had ever waved .</i>	0.52 ± 0.02	0.47 ± 0.02	0.00
subj_aux_agreement_question_vary_aux <i>was that man talking about those turtles ?</i> <i>*were that man talking about those turtles ?</i>	0.84 ± 0.01	0.84 ± 0.01	0.62
subj_aux_agreement_question_vary_subj <i>have the ladies yelled ?</i> <i>*have the lady yelled ?</i>	0.68 ± 0.01	0.63 ± 0.01	0.00
subj_aux_vary_subj <i>the doctor does n't talk .</i> <i>*the doctors does n't talk .</i>	0.78 ± 0.01	0.79 ± 0.01	0.29

Table 15: SCaMP_{plausible} results (5/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
subj_aux_vary_verb <i>that guy was fixing those toys .</i> <i>*that guy were fixing those toys .</i>	0.89 \pm 0.01	0.89 \pm 0.01	0.69
subj_aux_vary_verb_pp <i>the adult near the teachers is moving the pictures .</i> <i>*the adult near the teachers are moving the pictures .</i>	0.34 \pm 0.01	0.28 \pm 0.01	0.00
subj_aux_vary_verb_rc <i>this man who these people listen to has n't talked about this lady .</i> <i>*this man who these people listen to have n't talked about this lady .</i>	0.40 \pm 0.00	0.36 \pm 0.01	0.00
subj_verb_vary_verb <i>the king drives to these schools .</i> <i>*the king drive to these schools .</i>	0.67 \pm 0.01	0.63 \pm 0.01	0.00
svo_vos <i>these people have passed that hospital .</i> <i>*have passed that hospital these people .</i>	0.67 \pm 0.01	0.64 \pm 0.01	0.00
swapped_ditransitive_1 <i>she sent my doctor the hat .</i> <i>*my doctor sent the hat she .</i>	0.98 \pm 0.00	0.98 \pm 0.00	0.38
swapped_ditransitive_2 <i>we bought Kristen the toy .</i> <i>*Kristen bought we the toy .</i>	0.76 \pm 0.01	0.77 \pm 0.01	0.33
transitive <i>this dancer was making that bed .</i> <i>*this dancer was laughing that bed .</i>	0.95 \pm 0.00	0.96 \pm 0.00	0.00
wh_island <i>who does n't the girl learn she has taught ?</i> <i>*who does n't the girl learn who has taught ?</i>	0.66 \pm 0.01	0.62 \pm 0.01	0.00
wh_questions_object_gap <i>your mom did n't know the teacher who this lady can n't talk about .</i> <i>*your mom did n't know who this lady can n't talk about the teacher .</i>	0.70 \pm 0.01	0.63 \pm 0.01	0.00

Table 16: SCaMP_{plausible} results (6/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
wh_questions_object_gap_long_distance <i>those people may see the turtles that these children who these guys were n't talking about had heard .</i> <i>*those people may see what these children who these guys were n't talking about had heard the turtles .</i>	0.62 ± 0.02	0.55 ± 0.02	0.00
wh_questions_subject_gap <i>your teachers wo n't remember the men who have made the books .</i> <i>*your teachers wo n't remember who the men have made the books .</i>	0.94 ± 0.01	0.95 ± 0.00	0.00
wh_questions_subject_gap_long_distance <i>George had n't remembered that doctor who paints who could love this fork .</i> <i>*George had n't remembered who that doctor who paints could love this fork .</i>	0.96 ± 0.01	0.98 ± 0.00	0.00
wh_vs_that_no_gap <i>Rachel learned that that dentist likes the boy .</i> <i>*Rachel learned who that dentist likes the boy .</i>	0.98 ± 0.00	0.99 ± 0.00	0.01
wh_vs_that_no_gap_long_distance <i>that guy might remember that that girl who is running around the schools confused the people .</i> <i>*that guy might remember who that girl who is running around the schools confused the people .</i>	0.92 ± 0.01	0.93 ± 0.01	0.34
wh_vs_that_with_gap <i>my doctors forget what that driver made .</i> <i>*my doctors forget that that driver made .</i>	0.45 ± 0.01	0.47 ± 0.02	0.03
wh_vs_that_with_gap_long_distance <i>Andrew has learned what these guys who that driver did n't talk about stole .</i> <i>*Andrew has learned that these guys who that driver did n't talk about stole .</i>	0.18 ± 0.02	0.17 ± 0.02	0.80
overall	0.75 ± 0.00	0.76 ± 0.00	0.00

Table 17: SCaMP_{plausible} results (7/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
adjunct_island_implausible <i>what had this plate dropped without moving the women ?</i> <i>*what had this plate dropped the women without moving ?</i>	0.83 ± 0.01	0.84 ± 0.01	0.11
adv_order_implausible <i>that ring slowly worked with the plane .</i> <i>*that ring worked with slowly the plane .</i>	0.87 ± 0.01	0.91 ± 0.01	0.00
anaphor_gender_agreement_implausible <i>Jessica reads herself .</i> <i>*Jessica reads himself .</i>	0.92 ± 0.01	0.88 ± 0.01	0.00
anaphor_number_agreement_implausible <i>these candles helped themselves .</i> <i>*these candles helped himself .</i>	0.78 ± 0.01	0.75 ± 0.01	0.00
center_embedding_double_1_implausible <i>these bottles that doll that ship understands talks to laugh .</i> <i>*these bottles that doll that ship understands laugh .</i>	0.28 ± 0.01	0.28 ± 0.01	0.54
center_embedding_double_2_implausible <i>that rug the crayons the toys answer praise talks .</i> <i>*that rug the crayons the toys answer praise practice talks .</i>	0.96 ± 0.00	0.98 ± 0.00	0.00
center_embedding_single_1_implausible <i>this bowl those shoes talk to talks .</i> <i>*this bowl those shoes talks .</i>	0.31 ± 0.01	0.31 ± 0.01	0.30
center_embedding_single_2_implausible <i>that boat these shoes help exercises .</i> <i>*that boat these shoes help stretch exercises .</i>	0.96 ± 0.00	0.99 ± 0.00	0.00
complex_np_island_implausible <i>who do those schools who could n't care for those shoes understand ?</i> <i>*who do those shoes understand those schools who could n't care for ?</i>	0.56 ± 0.02	0.56 ± 0.02	0.93
coordinate_structure_constrain_object_extraction_implausible <i>who would n't the bed and the book listen to ?</i> <i>*who would n't the bed listen to and the book ?</i>	0.60 ± 0.03	0.62 ± 0.03	0.37

Table 18: SCaMP_{implausible} results (1/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
determiner_noun_agreement_1_implausible <i>the toys do n't teach those boxes .</i> <i>*the toys do n't teach those box .</i>	0.94 \pm 0.00	0.94 \pm 0.00	0.00
determiner_noun_agreement_2_implausible <i>the glass found that fork .</i> <i>*the glass found those fork .</i>	0.96 \pm 0.00	0.98 \pm 0.00	0.00
determiner_noun_agreement_adj_1_implausible <i>the chairs helped these nice boats .</i> <i>*the chairs helped these nice boat .</i>	0.90 \pm 0.00	0.89 \pm 0.00	0.01
determiner_noun_agreement_adj_2_implausible <i>the truck might praise that careful truck .</i> <i>*the truck might praise those careful truck .</i>	0.91 \pm 0.00	0.91 \pm 0.00	0.19
ellipsis_implausible <i>the rug is making two fine girls , and this car is making five .</i> <i>*the rug is making two girls , and this car is making five fine .</i>	0.60 \pm 0.01	0.61 \pm 0.01	0.29
green_ideas_implausible <i>kind cars stretch carefully .</i> <i>*stretch kind carefully cars .</i>	0.72 \pm 0.01	0.72 \pm 0.01	0.49
intransitive_implausible <i>those books smile .</i> <i>*those books hate .</i>	0.62 \pm 0.01	0.62 \pm 0.01	0.34
left_branch_island_simple_question_implausible <i>what spoon should these boxes know ?</i> <i>*what should these boxes know spoon ?</i>	0.77 \pm 0.01	0.79 \pm 0.01	0.01
matrix_question_npi_licensor_present_implausible <i>had that museum ever heard this picture ?</i> <i>*that museum had ever heard this picture .</i>	0.54 \pm 0.02	0.75 \pm 0.03	0.00
npi_present_implausible <i>the boat has often bothered the bicycle .</i> <i>*the boat has ever bothered the bicycle .</i>	0.48 \pm 0.01	0.49 \pm 0.01	0.08

Table 19: SCaMP_{implausible} results (2/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
only_npi_licensor_present_implausible <i>only those hats can ever drop that doctor .</i> <i>*even those hats can ever drop that doctor .</i>	0.57 \pm 0.05	0.72 \pm 0.07	0.00
only_npi_scope_implausible <i>only that shirt who would yell can ever understand these schools .</i> <i>*that shirt who would only yell can ever understand these schools .</i>	0.94 \pm 0.01	0.96 \pm 0.01	0.00
pp_order_implausible <i>this rocket visited that crayon on the floor .</i> <i>*this rocket visited on the floor that crayon .</i>	0.81 \pm 0.01	0.81 \pm 0.01	0.61
principle_A_c_command_implausible <i>this man who was saying your mother was writing himself .</i> <i>*this man who was saying your mother was writing herself .</i>	0.54 \pm 0.01	0.53 \pm 0.01	0.43
principle_A_domain_1_implausible <i>your mother was hoping that the doctors continue her .</i> <i>*your mother was hoping that the doctors continue herself .</i>	0.98 \pm 0.00	0.99 \pm 0.00	0.00
principle_A_domain_2_implausible <i>the doctors may say that your brother did n't read himself .</i> <i>*the doctors may say that your brother did n't read themselves .</i>	0.61 \pm 0.01	0.61 \pm 0.01	0.46
principle_A_domain_3_implausible <i>this man had said that this queen had said herself .</i> <i>*this queen had said that this man had said herself .</i>	0.50 \pm 0.00	0.51 \pm 0.00	0.00
recursion_intensifier_adj_long_implausible <i>the crayon is very very very very very very very very friendly .</i> <i>*the crayon reads very very very very very very very very friendly .</i>	0.62 \pm 0.03	0.72 \pm 0.03	0.00
recursion_intensifier_adj_medium_implausible <i>the plane was really really really really really strong .</i> <i>*the plane waves really really really really really strong .</i>	0.69 \pm 0.03	0.76 \pm 0.02	0.00
recursion_intensifier_adj_short_implausible <i>the bag is really really important .</i> <i>*the bag sings really really important .</i>	0.89 \pm 0.01	0.91 \pm 0.01	0.00

Table 20: SCaMP_{implausible} results (3/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
recursion_intensifier_adv_long_implausible <i>the couch cries really really really really really really really really really perfectly .</i> <i>*the couch is really really really really really really really really really perfectly .</i>	0.80 ± 0.03	0.89 ± 0.02	0.00
recursion_intensifier_adv_medium_implausible <i>the shoe talks very very very very very perfectly .</i> <i>*the shoe is very very very very very perfectly .</i>	0.91 ± 0.01	0.96 ± 0.01	0.00
recursion_intensifier_adv_short_implausible <i>the ring talks very very perfectly .</i> <i>*the ring was very very perfectly .</i>	0.89 ± 0.01	0.88 ± 0.01	0.69
recursion_poss_ditransitive_long_implausible <i>the ring sent your daughter 's teacher 's son 's cousin 's boss 's brother 's doctor 's daughter 's window the mirror .</i> <i>*the ring cleaned your daughter 's teacher 's son 's cousin 's boss 's brother 's doctor 's daughter 's window the mirror .</i>	0.51 ± 0.03	0.53 ± 0.04	0.48
recursion_poss_ditransitive_medium_implausible <i>the bowl bought your father 's boss 's teacher 's boss 's mom 's cousin 's book the shirt .</i> <i>*the bowl kept your father 's boss 's teacher 's boss 's mom 's cousin 's book the shirt .</i>	0.52 ± 0.04	0.53 ± 0.04	0.58
recursion_poss_ditransitive_short_implausible <i>the shoe sent my dentist 's mom 's shirt the mirror .</i> <i>*the shoe kept my dentist 's mom 's shirt the mirror .</i>	0.69 ± 0.02	0.73 ± 0.02	0.02
recursion_poss_transitive_long_implausible <i>the bed says that my father 's friend 's doctor 's mother 's teacher 's mom 's wife 's daughter 's glass is surprised .</i> <i>*the bed works with my father 's friend 's doctor 's mother 's teacher 's mom 's wife 's daughter 's glass is surprised .</i>	0.84 ± 0.03	0.90 ± 0.03	0.03
recursion_poss_transitive_medium_implausible <i>the plane thinks that your son 's grandmother 's cousin 's teacher 's button is kind .</i> <i>*the plane scares your son 's grandmother 's cousin 's teacher 's button is kind .</i>	0.97 ± 0.01	0.98 ± 0.01	0.03
recursion_poss_transitive_short_implausible <i>the doll thinks that your cup is angry .</i> <i>*the doll scares your cup is angry .</i>	1.00 ± 0.00	1.00 ± 0.00	0.03
recursion_pp_is_long_implausible <i>the boat on the couch on the couch in the school on the ground in the museum by the rug in the restaurant is wrong .</i> <i>*the boat was on the couch on the couch in the school on the ground in the museum by the rug in the restaurant is wrong .</i>	0.88 ± 0.02	0.91 ± 0.04	0.19

Table 21: SCaMP_{implausible} results (4/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
recursion_pp.is_medium_implausible <i>the boat on the floor on the floor in the school on the bed was kind .</i> <i>*the boat is on the floor on the floor in the school on the bed was kind .</i>	0.91 \pm 0.02	0.94 \pm 0.02	0.03
recursion_pp.is_short_implausible <i>the ring by the rocket was busy .</i> <i>*the ring is by the rocket was busy .</i>	0.97 \pm 0.01	0.99 \pm 0.00	0.00
recursion_pp.verb_long_implausible <i>the glass sitting in the school in the school in the hospital by the ring by the box on the ground by the car on the stool is friendly .</i> <i>*the glass sits in the school in the school in the hospital by the ring by the box on the ground by the car on the stool is friendly .</i>	0.49 \pm 0.06	0.52 \pm 0.09	0.66
recursion_pp.verb_medium_implausible <i>the glass sitting on the couch on the couch by the shirt in the kitchen is smart .</i> <i>*the glass sits on the couch on the couch by the shirt in the kitchen is smart .</i>	0.62 \pm 0.06	0.65 \pm 0.08	0.54
recursion_pp.verb_short_implausible <i>the bicycle sitting on the table is boring .</i> <i>*the bicycle sits on the table is boring .</i>	0.84 \pm 0.03	0.88 \pm 0.04	0.08
sentential_negation_npi.licensor_present_implausible <i>those toys did n't ever exercise .</i> <i>*those toys did probably ever exercise .</i>	1.00 \pm 0.00	1.00 \pm 0.00	0.74
sentential_negation_npi.scope_implausible <i>this toy who had waved can n't ever praise that spoon .</i> <i>*this toy who had n't waved can ever praise that spoon .</i>	0.50 \pm 0.02	0.47 \pm 0.02	0.04
subj_aux_agreement_question_vary_aux_implausible <i>does that hat lose ?</i> <i>*do that hat lose ?</i>	0.84 \pm 0.01	0.82 \pm 0.01	0.00
subj_aux_agreement_question_vary_subj_implausible <i>are n't the spoons waving ?</i> <i>*are n't the spoon waving ?</i>	0.68 \pm 0.00	0.64 \pm 0.01	0.00
subj_aux_vary_subj_implausible <i>the doll does n't visit this chair .</i> <i>*the dolls does n't visit this chair .</i>	0.73 \pm 0.01	0.72 \pm 0.01	0.02

Table 22: SCaMP_{implausible} results (5/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
subj_aux_vary_verb_implausible <i>the boxes were running .</i> <i>*the boxes was running .</i>	0.84 \pm 0.01	0.84 \pm 0.01	0.67
subj_aux_vary_verb_pp_implausible <i>that bicycle near the ladies was n't talking to that car .</i> <i>*that bicycle near the ladies were n't talking to that car .</i>	0.31 \pm 0.01	0.25 \pm 0.01	0.00
subj_aux_vary_verb_rc_implausible <i>these candles who work with that box do n't exercise .</i> <i>*these candles who work with that box does n't exercise .</i>	0.39 \pm 0.01	0.37 \pm 0.01	0.00
subj_verb_vary_verb_implausible <i>that ship reads .</i> <i>*that ship read .</i>	0.60 \pm 0.01	0.58 \pm 0.01	0.00
svo_vos_implausible <i>the hats might know this park .</i> <i>*might know this park the hats .</i>	0.43 \pm 0.01	0.39 \pm 0.01	0.00
swapped_ditransitive.1_implausible <i>they showed that bicycle the chair .</i> <i>*that bicycle showed the chair they .</i>	1.00 \pm 0.00	1.00 \pm 0.00	nan
swapped_ditransitive.2_implausible <i>he showed this bottle the shirts .</i> <i>*this bottle showed he the shirts .</i>	0.98 \pm 0.00	1.00 \pm 0.00	0.00
transitive_implausible <i>the plate likes the shirts .</i> <i>*the plate talks the shirts .</i>	0.89 \pm 0.00	0.91 \pm 0.00	0.00
wh_island_implausible <i>who had n't these candles realized they answer ?</i> <i>*who had n't these candles realized who answer ?</i>	0.50 \pm 0.02	0.50 \pm 0.02	0.81
wh_questions_object_gap_implausible <i>this rug had n't seen the plate that this museum would n't have .</i> <i>*this rug had n't seen what this museum would n't have the plate .</i>	0.78 \pm 0.02	0.74 \pm 0.01	0.00

Table 23: SCaMP_{implausible} results (6/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.

Category	Standard	Prior-trained	<i>p</i> -value
wh_questions_object_gap_long_distance_implausible <i>the cup wo n't see the guys who this museum who dropped those doctors would worry .</i> <i>*the cup wo n't see who this museum who dropped those doctors would worry the guys .</i>	0.72 ± 0.02	0.63 ± 0.02	0.00
wh_questions_subject_gap_implausible <i>these spoons have n't seen this patient who worried that glass .</i> <i>*these spoons have n't seen who this patient worried that glass .</i>	0.92 ± 0.01	0.95 ± 0.00	0.00
wh_questions_subject_gap_long_distance_implausible <i>those windows would see the teachers that were n't growing that heard the chair .</i> <i>*those windows would see what the teachers that were n't growing heard the chair .</i>	0.95 ± 0.01	0.98 ± 0.00	0.00
wh_vs_that_no_gap_implausible <i>this shirt learned that this shirt confused that crayon .</i> <i>*this shirt learned who this shirt confused that crayon .</i>	1.00 ± 0.00	1.00 ± 0.00	0.59
wh_vs_that_no_gap_long_distance_implausible <i>the window noticed that those hats who are n't fixing the girls could n't hate those bowls .</i> <i>*the window noticed who those hats who are n't fixing the girls could n't hate those bowls .</i>	0.97 ± 0.00	0.98 ± 0.00	0.05
wh_vs_that_with_gap_implausible <i>the buttons what n't remember what this couch has taken .</i> <i>*the buttons can n't remember that this couch has taken .</i>	0.29 ± 0.01	0.31 ± 0.02	0.03
wh_vs_that_with_gap_long_distance_implausible <i>this toy had realized what the bottles who are n't dropping that guy are n't keeping .</i> <i>*this toy had realized that the bottles who are n't dropping that guy are n't keeping .</i>	0.13 ± 0.01	0.13 ± 0.01	0.97
overall	0.73 ± 0.00	0.75 ± 0.00	0.00

Table 24: SCaMP_{implausible} results (7/7). For each model, the table shows the mean and a 95% confidence interval. For each category, there is one example of a minimal pair used to evaluate the category; a model is considered correct on an example if it assigns a higher probability to the sentence without an asterisk than the sentence with an asterisk.