Appendix A: A List of Stimuli in Definiteness Condition

Trial Type	Sentence 1	Sentence 2	Theme	Location 1	Location 2	Distractor
Indefinite trial	Pick up the calculator.	Now put it inside a box.	*small calculator*	*big box*	*small box*	*big container*
Indefinite trial	Pick up the clip.	Now put it inside a cup.	*small clip*	*big cup*	*small cup*	*big container*
Indefinite trial	Pick up the tape.	Now put it inside a box.	*small tape*	*big box*	*small box*	*big bowl*
Indefinite trial	Pick up the cup.	Now put it inside a box.	*small cup*	*big box*	*small box*	*big can*
Definite trial	Pick up the pan.	Now put it inside the bowl.	small pan	*big bowl*	*small bowl*	*big container*
Definite trial	Pick up the sponge.	Now put it inside the bag.	*small sponge*	*big bag-2*	*small-bag-2*	*big bowl*
Definite trial	Pick up the eggplant.	Now put it inside the bowl.	*small eggplant*	*big bowl*	*small bowl*	*big cup*
Definite trial	Pick up the bowl.	Now put it inside the container.	*small bowl*	*big container*	*small container*	*big box*
Filler trial 1	Pick up the can.	Now put it inside a bag.	*big can*	*big bag*	*small bag*	*big bowl*
Filler trial 1	Pick up the lock.	Now put it inside a can.	*big lock*	*big can*	*small can*	*big cup*
Filler trial 1	Pick up the pumpkin.	Now put it inside a container.	*big pumpkin*	*big container*	*small container*	*big cup*
Filler trial 1	Pick up the box.	Now put it inside a can.	*big box*	*big can*	*small can*	*big bag*
Filler trial 1	Pick up the bottle.	Now put it inside the cup.	*big bottle*	*big cup*	*small cup*	*big can*
Filler trial 1	Pick up the pig.	Now put it inside the cup.	*big pig*	*big cup*	*small cup*	*big bowl*
Filler trial 1	Pick up the post-it note.	Now put it inside the jar.	*big post-it note*	*big jar*	*small jar*	*big can*
Filler trial 1	Pick up the jar.	Now put it inside the container.	*big jar*	*big container*	*small container*	*big bag*

Filler trial 2	Pick up the potato.	Now put it inside the container.	big potato	big cup	small cup	big container
Filler trial 2	Pick up the duck.	Now put it inside the cup.	big duck	big box	small box	big cup
Filler trial 2	Pick up the strawberry.	Now put it inside the box.	big strawberry	big jar	small jar	big box
Filler trial 2	Pick up the cow.	Now put it inside the box.	big cow	big cup	small cup	big box
Filler trial 2	Pick up the squirrel.	Now put it inside the jar.	small squirrel	big can	smalll can	big jar
Filler trial 2	Pick up the orange.	Now put it inside the box.	small orange	big container	small container	big box
Filler trial 2	Pick up the sheep.	Now put it inside the can.	small sheep	big bowl	small bowl	big can
Filler trial 2	Pick up the mouse.	Now put it inside the cup.	small mouse	big can	small can	big cup

Note. There were two types of the filler trials. Filler trial 1 used a bigger theme so that only one of the two possible locations was available (e.g., a big pig does not fit in the smaller can. Filler trial 2 asked participants to put the theme inside the distractor location (e.g., "*Now put it inside the bowl*").

Appendix B: A List of Stimuli in Mass-Count Condition

Trial Type	Sentence 1	Sentence 2	Mass	Count	Theme	Theme
Quantifier	Pick up the camera.	Now put it on top of some jewelry.	jewelry	coins	camera	bicycle
Quantifier	Pick up the wallet.	Now put it on top of some rice.	rice	donuts	elephant	wallet
Quantifier	Pick up the watch.	Now put it on top of some butter.	butter	pandas	watch	cat
Quantifier	Pick up the banana.	Now put it on top of some ice.	ice	erasers	plate	banana
Numeral	Pick up the car.	Now put it on top of the two chewing gum.	chewing gum	batteries	car	horse
Numeral	Pick up the key.	Now put it on top of the two sand.	sand	flowers	chair	key
Numeral	Pick up the bomb.	Now put it on top of the two meat.	meat	straws	bomb	lipstick
Numeral	Pick up the turtle.	Now put it on top of the two popcorn.	popcorn	tigers	hot dog	turtle
Filler	Pick up the frog.	Now put it on top of the two sausages.	lettuce	sausages	frog	desk
Filler	Pick up the train.	Now put it on top of the two apples.	broccoli	apples	toothbrush	train
Filler	Pick up the tomato.	Now put it on top of the two pencils.	wood	pencils	tomato	skateboard
Filler	Pick up the ring.	Now put it on top of the two cookies.	corn	cookies	knife	ring
Filler	Pick up the laptop.	Now put it on top of some spoons.	toast	spoons	laptop	hat
Filler	Pick up the belt.	Now put it on top of some candles.	bacon	candles	cell phone	belt
Filler	Pick up the umbrella.	Now put it on top of some balls.	spinach	balls	umbrella	egg
Filler	Pick up the bear.	Now put it on top of some lemons.	asparagus	leomons	toilet paper	bear

Appendix C: Mass-count nouns with naturalness ratings from the norming study

In order to select count and mass nouns that were maximally distinguishable, we collected norming data from a separate group of 30 native English speakers, recruited through Amazon Mechanical Turk. Participants were asked to indicate how natural a word sounded in the following sentence: "*Please put three* ______ *here*." A score of 7 indicated '*most natural*,' while a score of 1 indicated '*least natural*.' From an initial list of 46 concrete nouns, we selected 16 count nouns with higher ratings (M = 5.95, SD = 0.32) and 16 mass nouns with lower ratings (M = 3.17, SD = 0.49). Average naturalness ratings of the count nouns were significantly greater than those of the mass nouns, $t(30)$ = 19.03, $p < .001$ (see the table below for item-specific ratings).

Count Noun	***M***	***SD***	**Mass Nouns**	***M***	***SD***
knife	5.17	1.92	sand	2.20	1.92
bell	5.59	1.96	jewelry	2.40	1.85
battery	5.63	2.06	wood	2.60	2.01
cat	5.66	1.90	rice	2.77	1.99
fork	5.83	2.02	ice	3.00	2.29
coin	5.87	1.93	spinach	3.00	1.91
stone	5.93	1.57	lettuce	3.13	1.87
clip	5.97	1.72	toast	3.17	1.88
candle	6.00	1.79	meat	3.20	1.68
book	6.07	1.93	chewing gum	3.40	2.01
carrot	6.13	1.65	butter	3.47	1.78
flower	6.13	1.73	broccoli	3.50	2.08
cookie	6.13	1.71	corn	3.50	2.01
pencil	6.23	1.67	bacon	3.73	2.22
apple	6.27	1.48	popcorn	3.73	2.21
lemon	6.53	0.85	asparagus	3.86	1.91

Note. The score 7 indicates '*a word sounds most natural*' and 1 indicates '*a word sounds least natural*' in the following sentence: "*Please put three* ______ *here*."

Appendix D: Scoring method SRT task

There were two probabilistic rules of sequence of dots: Sequences A and B. Sequence A (1–2–1–4–3–2–4–1–3–4–2–3) occurred with a probability of 85%, and Sequence B (3–2–3–4–1–2–4–3–1–4–2–1) occurred with a probability of 15% in one block. These sequences were comprised entirely of second-order conditionals (Reed & Johnson, 1994). A second-order conditional sequence refers to the sequence that is the third location of the triplet and is only determined by the combination of the previous two locations. For instance, Sequence A had a triplet (3-4-2), whereas Sequence B had another triplet (3-4-1). The first two locations (3-4) were equal between Sequence A and Sequence B, but the third location was different for Sequence A (2) than Sequence B (3). This way, the probability of co-occurrence of the third location with the pair (3-4) was 85% (Sequence A) and 15% (Sequence B).

The current study adopted the scoring method devised in Kaufman et al. (2010). First, the effect size (Cohen's *d*) was computed for the difference between mean RT for training trials and the mean RT for control trials in each block. Participants then received a point for each block in which learning was higher than the effect size × standard deviation for RT on each block.[1] As was found in Kaufman et al. (2010), the RTs in the training condition became consistently faster than those in the control condition from Block 3 to 8; the learning effect was not established at blocks 1-2, which were excluded from the analysis (see Kaufman et al., 2010 for the same decision). The scores across the last six blocks were summed up, resulting in the score

[1] Kaufman et al. used the grand effect size over the trials across all the blocks (Block 3-8). The current study found that this scoring method yielded lower reliability than the current scoring, which uses the effect size for each block, not the grand effect size. This may be because the effect sizes in each block differed, and using effect sizes for each block was more sensitive to measure learning effects.

ranging from 0 to a maximum of 6. According to Kaufman et al., using the average effect size is one way of setting a non-arbitrary criterion for learning effect, and the new scoring method obtained a higher reliability index than the conventional scoring method. Note that the more conventional scoring method (i.e., raw RT differences between the training and control conditions) showed a similar pattern of results but the new scoring system yielded stronger effects. According to the Kolmogorov-Smirnov test, the SRT score was not normally distributed ($ps < .05$). Data transformation did not effectively reduce the skewness for the SRT score; no transformation was applied.

Appendix E: Recognition Test

Task Design and Procedure. After the SRT task, participants also took a surprise recognition test. The test assessed whether participants became aware of the sequence patterns in the SRT task, i.e., whether they developed explicit knowledge about the sequence (Granena, 2013; Shanks & Johnstone, 1999). The presence (lack) of awareness for the sequence was assessed with an objective (RT) and a subjective measure (confidence ratings). In the recognition task, participants were told to respond to a dot of the triplet as quickly as possible and indicate whether they remember the triplet with a confidence rating. After every test item (i.e., triplet), participants rated their familiarity by giving a confidence rating on a 6-point scale ranging from 1 (I'm sure that this sequence was part of the test) to 6 (I'm sure that this sequence was not part of the test). The test consisted of half old (more familiar) and half new (less familiar) elements.

Following Granena (2013b), the 12 old triplets were constructed following second order conditionals in Sequence A (3–4–2, 3–1–2, 1–4–3, 2–4–1, 4–2–3, 1–2–1, 4–3–2, 4–1–3, 2–3–1, 2–1–4, 3–2–4, 1–3–4) and the 12 novel ones were constructed following second order conditionals in Sequence B (3–4–1, 3–1–4, 1–4–2, 2–4–3, 4–2–1, 1–2–4, 4–3–1, 4–1–2, 2–3–4, 2–1–3, 3–2–3, 1–3–2). These two sequences allowed us to examine to what extent the participants learned the second-order conditional information. The first two locations in every triplet were the same between the old (Sequence A) and new (Sequence B) triplets, but the third location was different (e.g., transition 3–4 was followed by location 2 in Sequence A and by location 1 in Sequence B).

Analysis. In the recognition test, the awareness of the SRT task was assessed in two ways: RTs and confidence ratings between the old and new triplets. First, faster RTs to the third dot in the old triplets than in the new triplets were expected because participants were trained to learn the second-order conditional information of Sequence A more than that of Sequence B. Second, the familiarity ratings were expected to be equal between the new and the old triplets, suggesting that participants cannot discriminate the learned sequence consciously. Evidence of poor recognition, but faster RTs, for segments of the old sequences was taken to suggest that the knowledge acquired during the training task produced behavioral effects before these effects are consciously attributed to the results of learning—implicit knowledge.

Results. In order to check explicit knowledge that did not influence the SRT performance, RTs and confidence ratings from the recognition test were compared between the new and old triplets. First, mean confidence ratings given to old and new sequences were 2.37 (*SD* = 0.79) and 2.45 (*SD* = 0.84). A paired-sample t-test showed no significant difference between the old and new sequences, $t(64) = 1.52$, $p = .13$, $d = 0.11$. Lack of conscious discrimination of old sequences from new sequences suggests that the SRT performance reflected implicit learning with little influence of explicit knowledge. Second, the mean RT on the third dot was 507 ms (*SD* = 152) and 540 ms (*SD* = 193) for the old and new sequences, respectively. A paired-sample t-test showed that the RT in the old sequence was significantly faster than in the new sequence, $t(64) = 2.54$, $p = .01$, $d = 0.18$. Faster RT in the old sequences than in the new sequences provides "a direct index of the possible influence of unconsciously applied perceptual-motor programs" (Shanks & Johnstone, 1999, p. 1446). This ensured, in combination with the confidence ratings results, that participants developed implicit knowledge of sequence with little influence from explicit knowledge. It endorses the validity of the SRT score as a measure for implicit learning aptitude.

Appendix F: Results from Data Analysis Including All Participants

The results presented here were from data analysis without excluding participants who reported to have noticed the target structure(s) in the visual-world task.

Descriptive statistics of sensitivity indices from the visual-world task (n = 65)

	M	*SD*	Min	Max	Possible Max
Def. + Mass-count	0.00	1.43	-4.20	3.34	-
Definiteness	-0.07	0.29	-0.73	0.73	1
Mass-count	0.18	0.30	-0.47	0.82	1

Correlations between Eye-tracking Scores and Aptitude Scores (n = 65)

	SRT	LLAMA F	MLAT_4
Combined	.29*	.01	.05
	(.01)	(.48)	(.36)
Definiteness	.36**	.01	-.10
	(.00)	(.97)	(.43)
Mass-count	.07	-.01	.14
	(.59)	(.95)	(.26)

Note. *P* values are indicated in brackets. * $p < .05$, ** $p < .01$

Multiple Regression Results

Composite of Definiteness and Mass-Count (n = 65)

	B	**SE**	***β***	***t***	***P***	**Partial-*r***
SRT	.07	.03	.29	2.35	.02	.29
LLAMA F	.02	.07	.03	.25	.81	.03
MLAT_4	.06	.23	.03	.27	.79	.03

Note. B and β indicate the unstandardized and standardized regression coefficients, respectively.

Definiteness (n = 65)

	B	**SE**	***β***	***t***	***P***	**Partial-*r***
SRT	.07	.02	.37	3.09	.00	.37
LLAMA F	.04	.05	.08	.67	.50	.09
MLAT_4	-.20	.19	-.13	-1.04	.30	-.13

Note. B and β indicate the unstandardized and standardized regression coefficients, respectively.

Mass-count (n = 65)

	B	**SE**	***β***	***t***	***P***	**Partial-*r***
SRT	.01	.03	.06	.48	.63	.06
LLAMA F	-.02	.06	-.04	-.31	.76	-.04
MLAT_4	.24	.21	.15	1.16	.25	.15

Note. B and β indicate the unstandardized and standardized regression coefficients, respectively.

References

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning, 63*(4), 665–703. doi:10.1111/lang.12018

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 585–594.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1435–1451.