Quality of early parent input predicts child vocabulary 3 years later

Erica A. Cartmill[a,1], Benjamin F. Armstrong III[b], Lila R. Gleitman[b,1], Susan Goldin-Meadow[a], Tamara N. Medina[c], and John C. Trueswell[b]

[a]Department of Psychology, University of Chicago, Chicago, IL 60637; [b]Department of Psychology, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA 19104; and [c]Department of Psychology, Drexel University, Philadelphia, PA 19104

Contributed by Lila R. Gleitman, May 20, 2013 (sent for review December 24, 2012)

Children vary greatly in the number of words they know when they enter school, a major factor influencing subsequent school and workplace success. This variability is partially explained by the differential quantity of parental speech to preschoolers. However, the contexts in which young learners hear new words are also likely to vary in referential transparency; that is, in how clearly word meaning can be inferred from the immediate extralinguistic context, an aspect of input quality. To examine this aspect, we asked 218 adult participants to guess 50 parents' words from (muted) videos of their interactions with their 14- to 18-mo-old children. We found systematic differences in how easily individual parents' words could be identified purely from this socio-visual context. Differences in this kind of input quality correlated with the size of the children's vocabulary 3 y later, even after controlling for differences in input quantity. Although input quantity differed as a function of socioeconomic status, input quality (as here measured) did not, suggesting that the quality of nonverbal cues to word meaning that parents offer to their children is an individual matter, widely distributed across the population of parents.

language acquisition | word learning | SES

Children's vocabularies vary greatly in size by the time they enter school (1, 2). Because preschool vocabulary is a major predictor of subsequent school success (3), this variability must be taken seriously and its sources understood. Some of this variability resides in the individual capacities and temperament that infants bring to the word learning task (4, 5). However, environmental influences are also bound to play instrumental roles. Accordingly, we examined the contextualized speech input parents provide to infants during the second year of life as a potential source of the massive vocabulary differences found at school entry.

It is already known that the sheer quantity of linguistic input is an important determinant of vocabulary size; overall, the more words children hear early in development, the larger their subsequent vocabularies. This relationship holds true both for types (different words) and tokens (number of words heard, including repetitions) (6, 7). These quantity differences are correlated with socioeconomic status (SES). Children from low SES homes are typically exposed to fewer words early in development (8, 9) and have smaller vocabularies at school entry than children from high SES homes (10).

Taken alone, the correlation of vocabulary size with amount of input is puzzling because as a general rule language learners do not seem to require a large number of exposures to a word to acquire its meaning (11). In experimental settings, for example, children have been shown to acquire and retain a new word heard only once or a very few times (12–14). The likelihood, then, is that certain exposures to a new word are especially informative, supporting secure and rapid inferences to meaning. For example, common sense insists that it will be easier to acquire the meaning of "zebra" in the visual presence of a zebra ("There goes a zebra!") than in its absence ("Let's visit the zebras in the zoo"). To that extent, an environment that maximizes this "here and nowness" of speech—its high quality, or referential transparency—might be expected to boost the rate of early word learning independent of the number of times a child hears each word.

Accordingly, we report here on the influence of such input-quality factors on word learning in infancy, asking: (*i*) how the referential transparency of input varies across families; (*ii*) how this variation impacts child vocabulary size 3 y later at school entry; and (*iii*) how this relationship interacts with SES and the quantity of linguistic input.

The first task in such an inquiry is to pin down a relevant description of input quality. Although, as just acknowledged, having an object in plain view when it is linguistically labeled must be useful, by itself this criterion is far from sufficient, because, in real life, every observed situation is replete with objects, events, properties, and relations. Thus, the learning problem becomes one of selection among many possible interpretations of the speaker's actual referential intention (15, 16).

Linguistic context is one well-known source of information that helps resolve this selection problem in older infants and toddlers who already have a considerable vocabulary and some appreciation of how words are put together syntactically in the language being learned (17–20). However, these potentially informative linguistic cues are largely inaccessible to the rank novices (14- to 18-mo-olds) whose learning environments we study here (21). In contrast, nonverbal cues (e.g., the presence and salience of a word's referent, whether it is being looked at, pointed to, or manipulated by the adult interlocutor) are available from earliest infancy and can, at times, simplify the task of inferring word meaning from situational context (22–25). Attempts have been made to investigate these nonverbal cues using detailed coding systems (26–28). Although such investigations have often been informative (for review, see ref. 29), the subtlety, variety, and sheer number of possible nonverbal cues make it difficult to decide a priori whether any particular contextualized utterance is in principle "good for learning" (26, 27). Rather than attempt to enumerate and classify these nonverbal environmental cues to a word's referent, we estimated the referential transparency of learning instances in infant-directed speech using a reliable and well-validated overall measure of referential transparency—how easily the meaning of a word can be inferred from nonlinguistic context alone—the so-called Human Simulation Paradigm (HSP) (16, 30).

In HSP, adult participants watch muted 40-s video clips ("vignettes") of actual parent-child interactions and try to guess the "mystery word" (indicated by a beep) that the parent uttered at a particular point in the video. The accuracy of these guesses is taken as a measure of how easily the meaning of the word can be inferred from situational context; that is, its quality as a learning

Author contributions: E.A.C., L.R.G., S.G.-M., and J.C.T. designed research; E.A.C., B.F.A., and T.N.M. performed research; E.A.C., B.F.A., T.N.M., and J.C.T. analyzed data; and E.A.C., L.R.G., S.G.-M., and J.C.T. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: ericacartmill@gmail.com or gleitman@psych.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309518110/-/DCSupplemental.

opportunity. Note that this measure of input quality is independent of the number of words a parent produces and thus is not confounded with input quantity. Prior studies show close agreement in adult and child responses in HSP, suggesting that they are sensitive to the same contextual properties (31, 32).

Evaluating Word-Learning Quality and Its Outcomes

Here we used HSP to test the following general prediction: Families who provide a greater proportion of high-quality word-learning opportunities early in childhood produce better vocabulary outcomes in their children. This prediction was tested in four steps: input sampling, quality assessment, quantity assessment, and language-attainment assessment.

Input Sampling. Fifty parent-child dyads from a stratified SES sample in the Chicago area were video recorded in their homes in whatever situations arose during two 90-min observation sessions at 14 and 18 mo of age (2, 33, 34). For each family, we randomly extracted 10 40-s vignettes, each showing the parent saying a different concrete noun (e.g., dog, ball) (*SI Text*) directed to the child.

Quality Assessment. Potential quality of each vignette was assessed via HSP: that is, showing the muted vignettes to adults ($n = 218$) who tried to guess the target word (indicated by a beep) that the parent uttered. Average accuracy was taken as the measure of the quality of that parent's input; a good learning opportunity occurs when the word's meaning can be readily inferred from its situational environment.

Quantity Assessment. We calculated each parent's average number (in tokens) of spoken words per minute during the two observation sessions at 14–18 mo.

Language-Attainment Assessment. A standardized vocabulary assessment [Peabody Picture Vocabulary Test (PPVT) (35)] was administered to the children at 54 mo. Scores served as the outcome measure at school entry, allowing us to ask how individual differences in quality and quantity of parent input early in development (14–18 mo) correlate with children's comprehension vocabulary 3 y later.

Results and Discussion

We found that the quality of socio-visual input to word meaning varied widely across the 50 parents, with parent HSP accuracy scores ranging from 5–38% (mean 22% ± 8%). Thus, some parents' speech to their offspring rarely contained highly informative contextual cues to meaning, whereas others' did so relatively often. Strikingly, this parent-input quality difference at child age 14–18 mo significantly correlated with the children's vocabularies at 54 mo (linear regression, $r^2 = 0.12$, $P = 0.014$) (Fig. 1*A*). [In all cases of linear regressions reported in this article, corresponding regressions were also computed removing any family identified as an "outlier" (operationalized as having a Cook's distance score greater than 0.15). Unless otherwise noted, these analyses yielded the same patterns of statistical significance.]

Quantity of speech input also varied widely across the 50 parents, with the number of words per minute produced during the 90-min videotaping sessions at ages 14 and 18 mo ranging from 6.3 to 97.0 words per minute (mean 37.8 ± 20.0). Furthermore, replicating previous work (6), this measure of quantity correlated with children's vocabularies at 54 mo (linear regression, $r^2 = 0.13$, $P = 0.011$).

The positive relationship between quality of input and later child vocabulary held even after statistically controlling for effects of quantity on this outcome measure [$r^2 = 0.22$, t(quality) = 2.40, P(quality) = 0.020] (Fig. 1*B*). When taken together (in a multiple regression), the measures of quality and quantity of early parent input at 14–18 mo accounted for 22% of the variance in children's

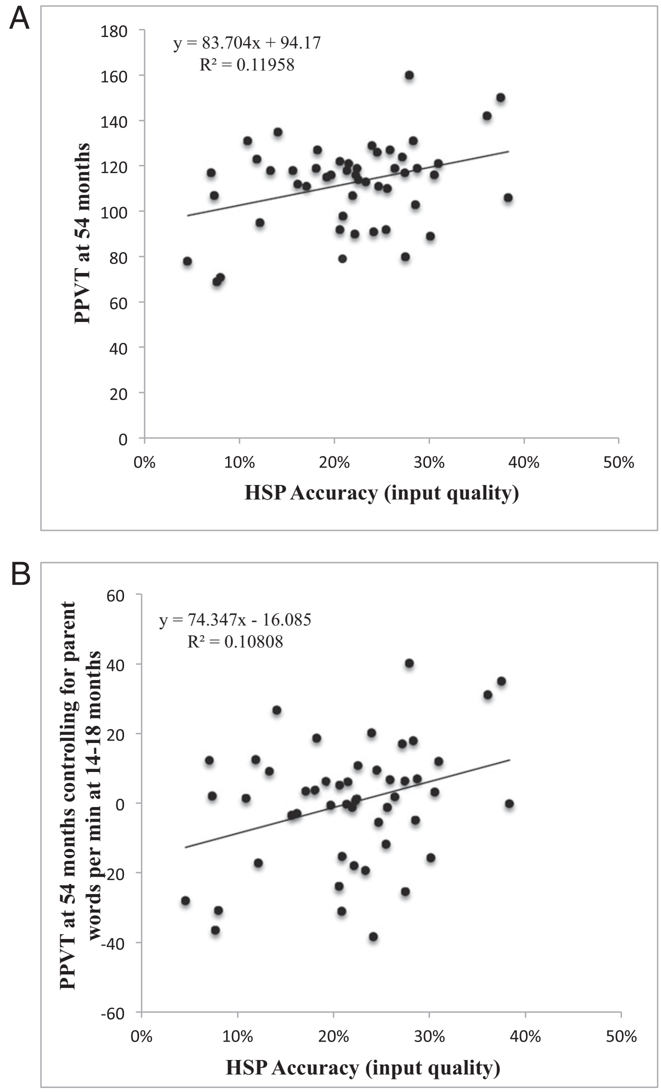

Fig. 1. Effect of quality of early input at 14–18 mo on child comprehension vocabulary at 54 mo. (*A*) Quality of word learning instances (average HSP accuracy per family) at 14–18 mo predicts child comprehension vocabulary (PPVT) at 54 mo. (*B*) This effect holds even after controlling for the quantity of early input (average parent words per minute at 14 and 18 mo). Each point represents a single family ($n = 50$).

vocabularies at 54 mo, a surprisingly strong relation given that 3 y, and presumably many life changes, intervened between assessments.

We next asked how quantity and quality combine to impact linguistic growth. First, quantity of parent input did not correlate with the HSP measure of input quality (Pearson correlation, $r = 0.11$ $P = 0.454$). That is, parents who talked more to their children did not, as a group, provide proportionally more or less high-quality word-learning instances. Second, quality and quantity did not interact with (or moderate) each other when predicting vocabulary outcome in a multiple regression that simultaneously included the main effects of quality, quantity and the interaction term ($P = 0.874$). That is, early quality and quantity accounted for different aspects of the variance found in the later vocabulary outcome measure.

However, parents who talk more are, by definition, offering their children more words, and the more words a child hears, the more likely it will be for that child to hear a particular word in a high-quality learning situation. That is, sheer parental talkativeness increases the likelihood that a child will encounter high-quality learning instances for particular words during natural parent-child conversational exchanges. To illustrate this effect, we generated an estimate of the total number (as opposed to the proportion) of high-quality learning instances a child receives by multiplying the HSP measure of input quality for each parent by that parent's input quantity measure. This number significantly correlated with child vocabulary at 54 mo (simple linear regression, $r^2 = 0.20$, $P = 0.001$; that is, the interaction term of quality times quantity reliably predicts vocabulary outcome on its own). Numerically, this is a stronger correlation than the correlation between quantity and child vocabulary at 54 mo (0.20 vs. 0.13), suggesting that the number of quality word-learning instances may matter more than the total number of words heard; however, these two correlation coefficients were not significantly different from each other in a comparison test of overlapping coefficients ($z = -1.26$, $P = 0.21$), which takes into account the fact that these two predictors are also highly correlated with each other ($r^2 = 0.72$, $P < 0.001$).

Did quality and quantity of early word-learning opportunities covary with the SES of our families? Primary caregiver education (range: <10–18+ y) and family income (range: <$7,500 to $100,000+) were combined to create a single SES variable (see *SI Text* for details). Replicating previous work (10), we found that SES correlated with the quantity of parent input ($r^2 = 0.15$, $P = 0.006$) (Fig. 2*A*). However, we found that SES did not correlate with our proportional (HSP) measure of quality ($r^2 = 0.005$, $P = 0.625$) (Fig. 2*B*). When both quality of parent input and family SES were considered as predictors of child vocabulary (in a multiple regression), the model accounted for 35% of the variance in vocabulary at 54 mo, and both quality and SES remained significant predictors [$r^2 = 0.35$, t(quality) = 2.64, P(quality) = 0.011]. Thus, parent SES and our measure of quality are not related to each other, and account for different aspects of the variance found in child vocabulary size at 54 mo.

Input to Naïve Learners. The advantage of studying input to word learning using videos of natural parent-child interactions is that it captures the real variability and complexity of the environments in which word learning naturally occurs. The drawback is that HSP participants have the opportunity to observe the child's reaction following the parent's utterance. If the child already knows the word the parent is saying and reacts accordingly, our HSP participants might be using that reaction to guess the mystery word rather than, or in addition to, using the utterance's concordance with objects and actions in view. For example, in a video where the target word was "nose," a parent might say: "Show me your nose." Vignettes in which the child already knows the word and points to her nose will be easier to guess correctly than videos in which the child does not know the word. Although we made extensive efforts to exclude such interactions from our vignette sample, it is still possible that child's prior knowledge of target words, rather than the informativity of parental use in context, could be driving HSP accuracy scores. If so, then the relationship between parent input quality and later child vocabulary could be nothing more than child vocabulary at 14–18 mo predicting later child vocabulary.

To address this problem, we assessed, on a child-by-child basis, whether there was evidence that the child was already producing any of the randomly sampled test words, using both observation of production (had the child ever spontaneously uttered this word during the 90-min video observation sessions) and a well-validated parental report measure of productive vocabulary [the MacArthur-Bates Communicative Development Inventory (CDI) (36)]. Discarding all vignettes for which there was evidence of prior production reduced both the number of vignettes per family (mean = 6.62 ± 1.78, range 3–10) and the number of families in the dataset ($n = 42$), but resulted in a set of vignettes ($n = 295$) that we could be reasonably certain contained target words not yet in the child's vocabulary.

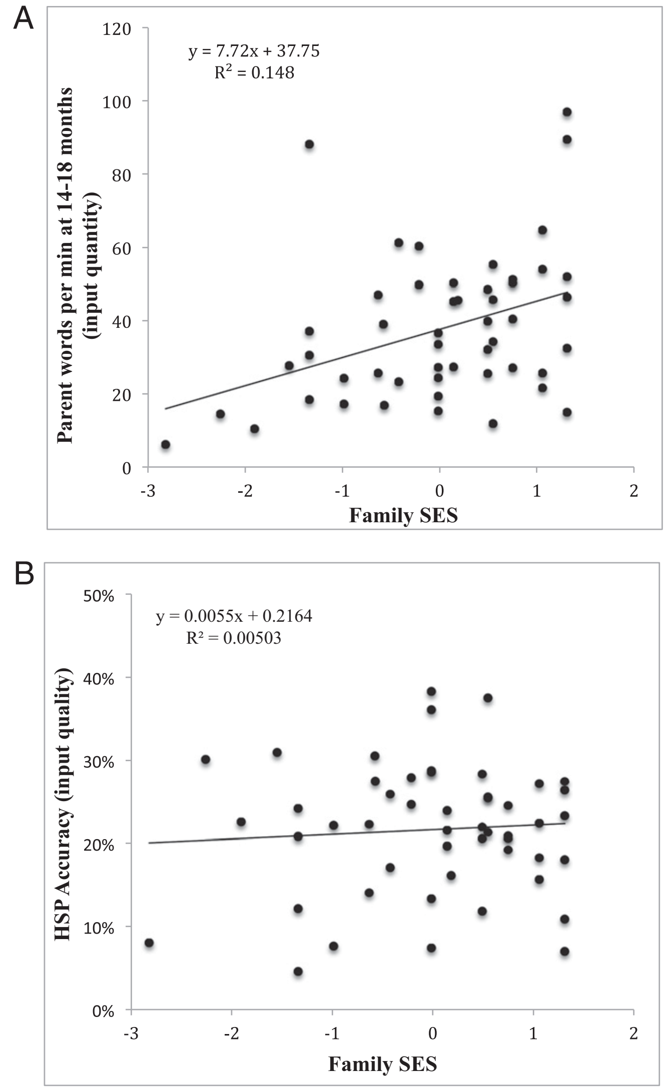

Fig. 2. Effect of SES on quantity and quality of parent input at 14–18 mo. (*A*) SES of family predicts average words per minute uttered by the child's parent during two 90-min recording sessions at 14 and 18 mo. (*B*) SES of family does not predict quality of parent input at 14 and 18 mo (accuracy measure calculated from HSP). Each point represents a single family ($n = 50$).

All analyses were then rerun on this reduced dataset of words not yet known by the child. If our reported correlation between quality and vocabulary outcome were actually a product of known words, we would expect this finding to be eliminated in this subset of unknown words. If, instead, the correlation reflects our ability to capture the quality of potential word-learning instances, the unknown subset should show a correlation similar to the correlation seen in the larger dataset. In all cases, the significant effect of parent input quality remained for unknown words. Specifically, average HSP accuracy for unknown words varied across the 42 parents and resembled the range of scores in the whole dataset (1–35%, mean 19% ± 9%). The effect of parent input quality (HSP accuracy) on children's vocabularies at 54 mo remained (linear regression, $r^2 = 0.11$, $P = 0.036$) (Fig. 3*A*). Conversely, parent input quality derived from videos of known words did not correlate with children's vocabularies (linear regression, $r^2 = 0.03$, $P = 0.306$).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

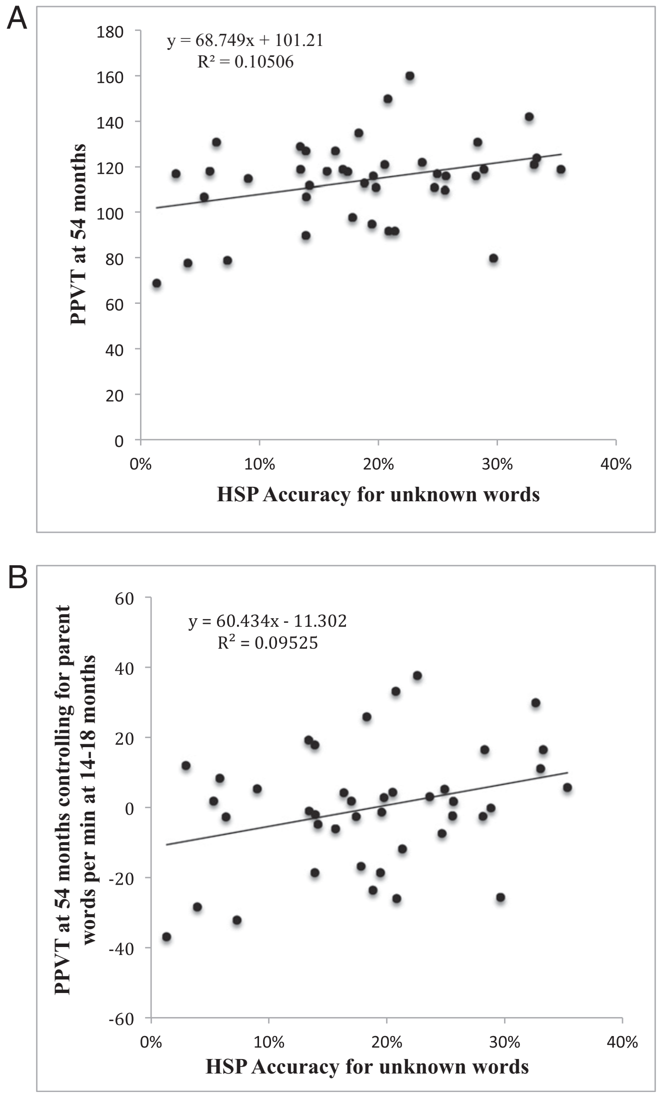

Fig. 3. Effect of quality of parent input for unknown words on child comprehension vocabulary at 54 mo. (*A*) Quality of unknown word learning instances (average HSP accuracy of vignettes in which child does not know the target word) at 14–18 mo predicts child comprehension vocabulary (PPVT) at 54 mo. (*B*) This effect holds even after controlling for the quantity of early input (average parent words per minute at 14 and 18 mo). Each point represents a single family (n = 42).

As before, the quality of parent input to naïve learners correlated with later child vocabulary even after controlling for any effects of the quantity of parent input [r^2= 0.23, t(quality) = 2.038, P(quality) = 0.048] (Fig. 3*B*). [In a corresponding multiple regression in which four families were removed as numerical outliers, the overall correlation remained significant (r^2 = 0.22, $P < 0.05$), but the quality coefficient was now marginally significant, t(quality) = 1.93, P(quality) = 0.062.] Also as before, quality of parent input did not correlate with quantity of parent input (Pearson correlation, r = 0.10 P = 0.520) or with family SES (Pearson correlation, r = 0.04 P = 0.827).

Thus, the strength of the relation between quality of parent input and later child vocabulary remained when restricting the assessment corpus to those items for which the child participants were less likely to have known the meanings of the words. We also controlled for children's total productive vocabulary at 18 mo (using CDI percentile scores) and found that parent input quality for unknown words continued to predict vocabulary at 54 mo, but CDI vocabulary score did not [r^2 = 0.14, t(quality) = 2.347, P(quality) = 0.024), t(vocabulary 18 mo) = −0.932, P(vocabulary 18 mo) = 0.357]. Together, these findings demonstrate that the effect of quality of parent input on future vocabulary size cannot be explained by differences in children's vocabularies at 14 and 18 mo.

Exploring Quality. An important next step in this line of research will be to identify the behavioral variables that contribute to high-quality input: that is, to identify which properties of the natural physical environment and which behaviors of a parent/speaker promote accurate reading of referential intent. For example, as one would certainly predict, vignettes in which the referent was visible (n = 178) at the beep were easier to guess than those in which the object was not visible (n = 117, 20% vs. 4% HSP accuracy). Laboratory experiments point to additional, social-pragmatic cues to guessing referential intent, including parent attention to the referent and parent gesture (deictic and nondeictic) (e.g., refs. 22 and 23). Notably, we replicate these findings in our own vignettes involving words not known by the child. In particular, vignettes in which parents were attending to the target object at the beep (n = 63) were significantly easier to guess than vignettes in which the object was present but the parent did not attend to it (n = 115): 40% vs. 9% HSP accuracy, $t(176) = -4.59$, $P < 0.001$.

Conclusions. Our findings provide significant and unique information about the conditions supporting vocabulary growth early in life. As quantified by a measure of interpretability in context (HSP), the coordination of word use with socio-visual aspects of parent-child interaction is a potent facilitator for the discovery of first word meanings. Grossly speaking, such an effect is self-evident. After all, how could novice word learning happen except by aligning the language sounds with their environmental contingencies? What is less expected is our first finding of great variability in individual parents' natural tendency to provide this contextual support regularly (ranging from 5 to 38% quality: that is, contextually informative, input), with increased informativity having a clear positive effect on vocabulary size by the time formal schooling begins 3 y later. Moreover, this variability in informativity is apparently an individual matter unrelated to SES, and thus seemingly uncorrelated with the more overt teaching styles and picture-book environments that are more prominent in high SES households (37). The positive effect of SES on vocabulary outcome is more likely related to the greater amount of talking by parents to their children in higher SES homes, which, in turn, increases the number of quality learning instances encountered overall.

As an important methodological note, the HSP measure of informativity has primarily been used with adult participant-judges, and these "ideal observer" findings have been assumed to be applicable to the child word-learning case as well. This assumption gains its a priori reasonableness from the finding that the kinds of words adults identify easily in the HSP paradigm are just the kinds of words infants learn early in language development (30). However, some commenters have questioned this inferential link between adult laboratory performance and child learning. The present study, which also used adult participant-judges, goes a long way toward alleviating any such provisos, putting HSP on a new and firmer explanatory footing. We have demonstrated here that adult HSP performance using videotapes of parental input to a sample of 14- to 18-mo-olds significantly predicts these children's vocabulary attainment 3 y later. In other words, the results of the adult HSP predicts word learning in the real world. Situational evidence supporting vocabulary growth seems to work in closely related ways, no matter the sophistication of the observer, and indifferently to whether they are acquiring a first or a second language (38, 39).

We can now return to the apparent paradox described in our introductory remarks: On the one hand, child word learning appears to be amazingly rapid during a period extending from infancy at least throughout late childhood. Indeed, the sheer math of vocabulary acquisition (mean receptive vocabulary of ∼12,000 words by the sixth birthday, 23, 40) is understandable, especially if word learning typically requires only one or a very few exposures to words in context for the meaning to be acquired. Recent experimental evidence supports a model [termed "propose but verify" (31, 41)] that comports with these findings by suggesting that learners form a single conjecture about word meaning given a context. Here "context" can be either the immediate situational environment or the immediate linguistic environment for the word [the latter becomes increasingly effective as the child acquires and builds grammatical knowledge (30)]. However, this hypothetical procedure, operating summarily on very little input data, must have a systematic way of avoiding false mappings that may seem plausible in any adventitiously observed situation. The learning procedure succeeds by a species of filtering: it discards or ignores low-quality encounters that happen along the way, thus preventing them from entering into the search for meaning. However, if this fast-mapping procedure is the one that best characterizes child learners, why should the sheer quantity of words children hear correlate so well with vocabulary attainment [as previously reported (6, 7) and as replicated herein]? This frequency-sensitivity finding might suggest a relatively "slow-mapping" learning procedure during which children (indiscriminately) aggregate across both more and less informative encounters with a word, eventually identifying the correct meaning probabilistically on a best-fit basis [as implemented in, for example, simple associationist models (42)]. There is, in fact, suggestive evidence for this kind of cumulative procedure, not only in the effects of frequency that we and others have reported, but also in the generalizations children make on the basis of frequent contexts [e.g., young children often insist that cousins are necessarily children and that uncles are adults (43)].

The present results offer a way to reconcile these positions. As we have documented herein, vocabulary attainment is significantly correlated with quality input, where "quality" is defined as referential transparency (operationalized as HSP score), an effect that is independent of "quantity" (and independent of SES as well). This result comports with the propose-but-verify fast-mapping procedure that makes conjectures only in the presence of high-quality information. However, arguably this procedure could lead to a frequency effect on vocabulary attainment all of the same, simply because frequency probabilistically increases opportunities for encountering highly informative learning instances: the only kind of instance that, when coupled with later confirmatory instances, pushes the child's vocabulary forward.

Summarizing, the present study is consistent with the position that words are learned via a relatively sudden, determinative, and insightful procedure, rather than by brute-force statistical machinery that is accumulating and cross-tabulating all observed instances and utterances across extended periods of time. Still, this fast-mapping machinery, just like statistical-learning machinery, entails an advantage for the more talkative over the more taciturn learning environments within a family and, derivatively, a correlation between SES and vocabulary attainment during the preschool years (because high SES families, as a group, provide higher quantity of input speech). However, the deeper point is that the environment that provides quality input supports efficient learning and that variations in quality, as we have defined and documented it, are observed across the SES spectrum.

In conclusion, the present findings dovetail nicely with laboratory studies of early language learning at the phonetic and word level, showing that an infant's ability to pick up on information in the social context is strongly linked to learning a novel natural language (22–25). The present study shows that this social information—quality encounters in which interpretation is transparent to socio-visual inspection—is delivered to varying degrees by parents in natural settings, and that its presence predicts later language skills.

Materials and Methods

Parent-Child Pairs. Fifty parent-child pairs were included in the analyses. All parents gave written informed consent to participate in the longitudinal study and for their videos to be used in additional research. Children were typically developing (27 males, 23 females). The pairs were from a larger sample of 63 families participating in a longitudinal study of language development (33, 34); see *SI Text* for details. Families were chosen to be representative of the ethnicity and income ranges of the greater Chicago area. All children were raised monolingual English speakers. Families were visited in their homes every 4 mo from child age 14 to 58 mo, and were video recorded for 90 min at each visit. During visits, families engaged in their normal daily activities, ranging from book reading and puzzle play to meals and bathing. Selection of families for present analyses ($n = 50$) was based on the following exclusion criteria (*SI Text*): parents did not permit videos to be used in future research ($n = 6$); the parent was not a native-English speaker ($n = 1$); the child did not take the PPVT test at 54 mo ($n = 6$).

The following measures were taken of each family (*SI Text*): (*i*) SES as determined by a principle component analysis (33) combining the education level of the primary caregiver (range: 10–18 y; mean = 15.96, SD = 2.16) and family annual income (range: under $7,500 to over $100,000; mean = $64,000, SD = $30,000); (*ii*) Quantity of parent's linguistic input was the average number of words per minute in child-directed speech produced by the parent at the 14- and 18-mo visits; (*iii*) Child vocabulary outcome was measured using the PPVT at 54 mo (35). We used the PPVT scores from 54 mo because they reflect child vocabulary just before school entry.

Vignette Selection for HSP. A total of 560 40-s muted videos (vignettes) served as target stimuli in the HSP study below. These vignettes came exclusively from the 14- and 18-mo-old visits. Each vignette was an example of a parent uttering one of the 40 most common concrete nouns in the transcript sample, uttered usually within a sentence context (e.g., *Can you give me the book?*). Vignettes were aligned so that ∼30 s into the video, the parent uttered the target word (at which point a beep was inserted). If the parent uttered the target word more than once during the 40-s vignette, each instance of the target word was replaced by a beep. We considered this local repetition of target words to be a feature of the quality of the context and thus allowed it to vary naturally. Previous studies found this duration to be sufficient to understand the gist at the moment the target word was uttered (31). The *SI Text* describes in detail our vignette selection criteria, but in brief, 10 vignettes were selected from each of 56 participating families. (This sample included the six families later excluded for lack of a PPVT test score at 54 mo.) Five filler videos were also selected from each family: that is, 280 fillers. Filler words consisted of verbs, adjectives, quantifiers, or nouns that were not easily visualized, so as to prevent participants from only guessing concrete nouns.

HSP Experimental Design. Because no single HSP participant could reasonably view and respond to all 840 40-s vignettes (9.33 h of video in total), vignettes were split into 15 experimental lists such that each list had no more than one vignette from each family and had no more than four examples of the same word.

HSP Participants. Participants ($n = 218$) were randomly assigned one of the 15 lists consisting of 56 vignettes (including both target and filler words). Participants were undergraduate students (145 female, 73 male) enrolled at the University of Pennsylvania ($n = 159$) or La Salle University ($n = 59$) in Philadelphia. All were native English speakers, coming from relatively diverse SES backgrounds (see *SI Text* for details). The protocol was approved by the Institutional Review Boards at the University of Chicago, the University of Pennsylvania, and La Salle University. Participants provided informed consent and received course credit or payment for participating.

HSP Procedure. After viewing a vignette, participants guessed the "mystery" word for that vignette before viewing the next. Participants were tested individually or in groups, ranging from one to six people. Video was projected on a wall or screen and participants recorded their guesses on paper. See *SI Text* for details.

Analyses. Participant guesses were scored as correct if they were identical to the target word. Abbreviations and plurals were also counted as correct (e.g., phone or phones for telephone), but words that altered the meaning of the root word were not (see *SI Text* for details, and ref. 31). Participants' responses were used to calculate the guessing accuracy (average number of correct guesses) for each video clip. The guessing accuracies of a parent's 10 target videos were then averaged to create an average HSP accuracy for that parent. This average was used as the measure of quality of input for each family.

We used linear regressions to examine the relationships between: (*i*) the quantity of parent input, (*ii*) the quality of parent input, (*iii*) child vocabulary at 54 mo, and (*iv*) family SES. In particular, we analyzed the ability of input quality and quantity to predict child vocabulary at 54 mo.

For analyses involving only those target words not yet known by the child, child knowledge of HSP target words was determined by combining the words the child actually produced during the two 90-min observation sessions with the words parents reported the child produced at 14 and 18 mo. Recorded production was measured during the 90-min video observation sessions at 14 and 18 mo (i.e., a child spontaneously produced a target word during that time). Parent report of production was determined through the MacArthur-Bates CDI (36), in which parents marked any words they believed their child had said. The CDI was given to parents at every observation session, but not all parents returned the CDI at each session. Only vignettes of parents who completed the CDI at 18 mo were included in the unknown word dataset. If either observation or parent report indicated that the child produced an HSP target word, it was deemed known. These "known-word vignettes" were excluded from analyses where our focus was on parent input to naïve learners.

ACKNOWLEDGMENTS. We thank the participating families for sharing their children's language development with us; all the research assistants who collected and transcribed the data; A. Hafri for programming and data collection; M. L. Rowe, J. G. Foster, and S. Raudenbush for methodological advice; and Y. Lin, K. Schonwald, and J. Voigt for administrative and technical assistance. This research was supported by National Institute of Child Health and Human Development Grants P01HD40605 (to S.G.-M.) and R01-HD-037507-11S1 (to L.R.G. and J.C.T.).

1. Fenson L, et al. (1994) Variability in early communicative development. *Monogr Soc Res Child Dev* 59(5):1–173, discussion 174–185.
2. Rowe ML, Raudenbush SW, Goldin-Meadow S (2012) The pace of vocabulary growth helps predict later vocabulary skill. *Child Dev* 83(2):508–525.
3. Anderson RC, Freebody P (1981) Vocabulary knowledge. *Reading Comprehension and Education*, ed Guthrie JT (International Reading Association, Newark, DE), pp 77–117.
4. Stromswold K (2001) The heritability of language: A review and metaanalysis of twin, adoption and linkage studies. *Language* 77(4):647–723.
5. Tsao FM, Liu HM, Kuhl PK (2004) Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Dev* 75(4): 1067–1084.
6. Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T (1991) Early vocabulary growth: Relation to language input and gender. *Dev Psychol* 27(2):236–248.
7. Hurtado N, Marchman VA, Fernald A (2008) Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Dev Sci* 11(6):F31–F39.
8. Hart B, Risley TR (1995) *Meaningful Differences in the Everyday Experiences of Young American Children* (Brookes, Baltimore, MD).
9. Hoff-Ginsberg E (1991) Mother-child conversation in different social classes and communicative settings. *Child Dev* 62(4):782–796.
10. Hoff E (2003) The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Dev* 74(5):1368–1378.
11. Carey S (1978) The child as word learner. *Linguistic Theory and Psychological Reality*, eds Halle M, Bresnan J, Miller GA (MIT Press, Cambridge, MA), pp 264–293.
12. Spiegel C, Halberda J (2011) Rapid fast-mapping abilities in 2-year-olds. *J Exp Child Psychol* 109(1):132–140.
13. Nappa R, Wessel A, McEldoon KL, Gleitman LR, Trueswell JC (2009) Use of speaker's gaze and syntax in verb learning. *Lang Learn Dev* 5(4):203–234.
14. Heibeck TH, Markman EM (1987) Word learning in children: An examination of fast mapping. *Child Dev* 58(4):1021–1034.
15. Fodor JA (1981) The present status of the innateness controversy. *Representations*, ed Fodor JA (MIT Press, Cambridge, MA), pp 257–316.
16. Gleitman LR (1990) The structural sources of word meaning. *Lang Acquis* 1(1):3–55.
17. Naigles L (1990) Children use syntax to learn verb meanings. *J Child Lang* 17(2): 357–374.
18. Gertner Y, Fisher C, Eisengart J (2006) Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychol Sci* 17(8):684–691.
19. Gleitman LR, Cassidy K, Nappa R, Papafragou A, Trueswell JC (2005) Hard Words. *Lang Learn Dev* 1(1):23–64.
20. Hoff E, Naigles L (2002) How children use input to acquire a lexicon. *Child Dev* 73(2): 418–433.
21. Bruner J (1974) From communication to language—A psychological perspective. *Cognition* 3(3):255–287.
22. Baldwin DA (1991) Infants' contribution to the achievement of joint reference. *Child Dev* 62(5):875–890.
23. Bloom P (2000) *How Children Learn the Meanings of Words* (MIT Press, Cambridge, MA).
24. Kuhl PK (2007) Is speech learning 'gated' by the social brain? *Dev Sci* 10(1):110–120.
25. Brooks R, Meltzoff AN (2008) Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *J Child Lang* 35(1):207–220.
26. Allen SEM (2000) A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics* 38(3):483–521.
27. Frank MC, Tenenbaum JB, Fernald A (2013) Social and discourse contributions to the determination of reference in cross-situational word learning. *Lang Learn Dev* 9(1): 1–24.
28. Laakso ML, Poikkeus AM, Katajamäki J, Lyytinen P (1999) Early intentional communication as a predictor of language development in young toddlers. *First Lang* 19(56): 207–231.
29. Hoff E (2006) How social contexts support and shape language development. *Dev Rev* 26(1):55.
30. Gillette J, Gleitman H, Gleitman L, Lederer A (1999) Human simulations of vocabulary learning. *Cognition* 73(2):135–176.
31. Medina TN, Snedeker J, Trueswell JC, Gleitman LR (2011) How words can and cannot be learned by observation. *Proc Natl Acad Sci USA* 108(22):9014–9019.
32. Piccin TB, Waxman SR (2007) Why nouns trump verbs in word learning: New evidence from children and adults in the Human Simulation Paradigm. *Lang Learn Dev* 3(4): 295–323.
33. Rowe ML, Goldin-Meadow S (2009) Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* 323(5916):951–953.
34. Rowe ML, Levine SC, Fisher JA, Goldin-Meadow S (2009) Does linguistic input play the same role in language learning for children with and without early brain injury? *Dev Psychol* 45(1):90–102.
35. Dunn LM, Dunn LM (1997) *Peabody Picture Vocabulary Test* (American Guidance Service, Circle Pines, MN), 3rd Ed.
36. Fenson L, et al. (1993) *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (Brookes, Baltimore, MD).
37. Fletcher KL, Reese E (2005) Picture book reading with young children: A conceptual framework. *Dev Rev* 25(1):64–103.
38. Snedeker J, Geren J, Shafto CL (2007) Starting over: International adoption as a natural experiment in language development. *Psychol Sci* 18(1):79–87.
39. Johnson JS, Newport EL (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognit Psychol* 21(1):60–99.
40. Anglin JM (1993) Vocabulary development: A morphological analysis. *Monogr Soc Res Child Dev* 58(10, Serial No. 238):1–186.
41. Trueswell JC, Medina TN, Hafri A, Gleitman LR (2013) Propose but verify: Fast mapping meets cross-situational word learning. *Cognit Psychol* 66(1):126–156.
42. Yu C, Smith LB (2007) Rapid word learning under uncertainty via cross-situational statistics. *Psychol Sci* 18(5):414–420.
43. Keil FC (1989) *Concepts, Kinds and Cognitive Development* (Bradford Books, Cambridge, MA).

Supporting Information

Cartmill et al. 10.1073/pnas.1309518110

SI Text

Demographics of Families in the Video Corpus. Videotapes from 50 parent-child pairs were included in the analyses for the study. The children in the sample were typically developing and roughly split between male and female (27 males, 23 females). Parent-child pairs were selected from a larger sample of 63 families used in a longitudinal study of language development (1, 2). The 63 families in the longitudinal study were chosen to be representative of the ethnicity and income ranges of the greater Chicago area. In the subset of 50 families included in the analyses we report, 36 of the children were White, 9 were African American, and 5 were Multiracial. Additionally, 5 reported their ethnicity as Hispanic and 45 as non-Hispanic. All children were being raised as monolingual English speakers. The parent in each pair was the adult who identified as the primary-caregiver for the child (49 mothers, 1 father). In cases where parents reported sharing caregiver duties, the parent that contributed the most child-directed speech in videotapes taken at 14 and 18 mo was chosen to represent parent input in that family.

As part of the longitudinal study, families were visited in their homes every 4 mo from child age 14 to 58 mo, and were videotaped for 90 min at each visit. During the visits, families were asked to engage in their normal daily activities. The videotapes thus included a wide range of activities, ranging from book reading and puzzle play to meals and bathing. The videotapes focused on the children and their immediate environment. Primary caregivers were usually present; the presence of siblings and other adults varied across families and visits.

The parent-child pairs chosen for the present study were those who had given permission for their videotapes to be used in further research. One family that met these criteria was further excluded because the caregiving parent was not a native English speaker and used some object labels in atypical ways (e.g., "car" for a fire truck). We thought this would unfairly bias naïve adult guessers against accurately guessing the words in that family's videotapes. These criteria resulted in a set of 56 parent-child pairs. Vignettes from these 56 pairs were experimentally tested in the Human Simulation Paradigm (HSP), but six of those families were later excluded from analysis because the children did not take the vocabulary test given at 54 mo. The experimental design presents details on the larger number of 56 families because the experiment was run with vignettes from all families. However, all other methods sections report data from only the 50 families used in analysis.

Socioeconomic status (SES) was calculated for each of the 50 families using the education level of the primary caregiver and family annual income. Families with high SES scores have a high annual income and the primary caregiver has a high level of education. Education ranged from 10 to 18 y, where 12 y was equivalent to completing high school or GED (mean = 15.96, SD = 2.16). Income ranged from less than $7,500 to more than $100,000 per year (mean = $64,000, SD = $30,000). Education and annual income were positively correlated ($r = 0.37$, $P = 0.009$). We combined education and income into one variable (SES) using principal component analysis following the procedure used in ref. 1. The first component accounted for 68% of the original variance.

Language Measures. The quantity of parent input was measured as the average number of words per minute in child-directed speech produced by the parent at the 14- and 18-mo visits. We used words per minute rather than total words as our measure of input because transcript length varied from 56 min to 101 min (mean = 87.91, SD = 5.24). The number of words per minute was calculated for the two time points separately and then averaged. Only child-directed parental speech was transcribed from the videotape. Speech to other children (usually siblings) was also included. In the transcription, all dictionary words, onomatopoeic words (e.g., "meow") and evaluative sounds (e.g., "uh-oh") were coded as words.

Child vocabulary was measured using the Peabody Picture Vocabulary Test (PPVT) (3) at 54 mo. The test was administered several times during the original longitudinal study (at 30, 42, and 54 mo, and then in kindergarten and second grade). Following other studies of these children (1), we used the PPVT scores from 54 mo because they reflect child comprehension vocabulary just before school entry, thus maximizing the period during which early parent input might have had an effect and minimizing the effect attributable to school input. A parent-report measure of productive vocabulary [the MacArthur-Bates Communicative Development Inventory (CDI) (4)] was used to measure children's vocabularies at 18-mo. Percentile CDI scores were used to control for early differences in child vocabulary size.

Child knowledge of target words was determined by combining the words children actually produced during the videotaped observations with the words parents reported that the children produced at 14 and 18 mo. Observed production was measured during the 90-min video observation sessions at 14 and 18 mo (i.e., a child spontaneously produced a target word during that time). Parent report of production was determined through the MacArthur-Bates CDI (4). Parents were asked to mark any words that they had heard their child say. All target words except for one ("step") were items on the CDI. The CDI was given to parents at every observation session, but not all parents returned the CDI at each session. We compared the ages of observed and reported production of target words to the ages at which the vignettes were recorded to identify those vignettes in which the child had already acquired the target word. These "known-word vignettes" were excluded from the second half of our analysis where we wanted to focus on parent input to naïve learners.

Stimulus Construction: HSP. Transcriptions of the 90-min videotapes were used to identify the most frequently occurring words across the entire corpus at the 14- and 18-mo videotapes. Words had to be used by at least six different parents and were ranked according to overall frequency. Two lists were constructed from the list of most frequent words: target words (concrete nouns) and filler words (other types of words included to keep participants from always guessing nouns).

Target words were frequent concrete nouns (e.g., book, car), which are among children's first words (4, 5) and have a higher probability than other words of being guessed from visual input alone (6). Potentially ambiguous nouns that were sometimes used as verbs (e.g., kiss, fish) were retained in the target list regardless of whether a parent used them as a noun or a verb to preserve the ambiguity present in the natural input. Filler items were verbs, adjectives, quantifiers, or nouns that were not easily visualized (e.g., job, today, music). The filler list was also constructed from the most frequent words, but items were selected to maximize the number of different types of filler words so that participants would not begin guessing only concrete nouns.

To construct a stimulus set of video clips ("vignettes") of parent input from each family, we selected 10 target words and 5 filler words for each parent from the frequency-ranked lists of potential target and filler words. We constructed a list of target and filler words for each parent by starting with the most frequent word on the target and filler lists and working our way down until we had 10 target and 5 filler words said by that parent. All target items were within the 115 most frequent concrete nouns in our corpus of child-directed speech at 14 and 18 mo. For each of a parent's chosen words, we randomly selected one instance from the parent's transcripts taken at child age 14 and 18 mo, examined the videotape for that instance, and used the criteria described in the next section to determine whether the videotape should be included in our stimulus materials. If the videotape did not meet our criteria, we randomly selected another instance of that target word from the parent's transcripts and repeated the procedure. If the videotapes for all of the instances of a particular target word that a parent produced failed to meet our criteria, we moved on to the next most frequent target word on that parent's list. This procedure continued until we had vignettes of 10 target items for each parent. We used the same procedure to select five filler items for each parent.

Forty-second-long vignettes of each parent's target and filler words were created by clipping the 30 s before and the 10 s after the selected instance of the word. Vignettes were not included in the stimulus set if they had any of the following properties (see ref. 7): (*i*) the referent of the target word was visible to the child, but could not be seen or identified on the video; (*ii*) it was possible to read the parent's lips or read the target word printed in a book or picture; (*iii*) there was insufficient lighting to clearly view the scene and it could not be corrected with video enhancement; (*iv*) the videotaping process added cues to meaning, such as panning to or zooming in on the referent; (*v*) more than one parent used the target word during the 40 s. The first of these criteria stems from our desire to give experimental participants access to all of the cues that the child had; in other words, to make the experimental task as similar as possible to the child's original learning environment. Thus, if the child did not have visual access to the referent in the actual event, the vignette was included as part of the stimulus set. For example, if the mother asked whether the child wanted to go to the pool to play in the water, the vignette was included in the stimulus set because, although the participant could not see water, neither could the child. Applying these criteria resulted in a total corpus of 560 target and 280 filler vignettes from our original set of 56 families. The set of vignettes from the 50 families included in the final analysis included 40 different target words and 16 different filler words (see Table S1 for a list of these words). Contrast, grain, and brightness needed to be adjusted on a number of vignettes to compensate for variation in lighting quality.

The audio of the vignettes was muted and each occurrence of the target or filler word was replaced by a beep occurring exactly when the parent had uttered the word. This procedure resulted in all of the vignettes having a beep at ≈30 s. For vignettes where the parent uttered the word more than once during the 40 s, additional beeps were included, each at the particular moment when the word had been uttered. Beeps per vignette (for the 50 families in the final dataset) ranged from 1 to 11 (mean = 2.14, SD = 1.55). We allowed the number of beeps in the vignettes to vary to both reflect natural individual differences in how often parents used particular words and to avoid biasing participants against guessing the correct word (e.g., if a parent was focusing on, and talking about, a teddy bear for an extended period, but there was only a single beep, participants might assume that the word could not be "bear"). We consider local repetitions of the target word to be a feature of the quality of the context and thus did not attempt to account for this variation during analysis.

Experimental Design. Vignettes were arranged into 15 lists such that each list had no more than one vignette from each family. The lists had no more than four examples of the same word (all from different families) and examples of the same word were separated by at least two intervening vignettes.

Experimental Protocol. Participants ($n = 218$) were randomly assigned one of the 15 lists consisting of 56 vignettes (one from each family in the full dataset). After viewing a vignette, participants guessed the "mystery" word for that vignette before viewing the next. Participants were tested individually or in groups, ranging from one to six people. The video was projected on a wall or screen and participants recorded their guesses on paper. The audio containing the beeps was played via external computer speakers. An experimenter was present during testing to control the video and monitor participants' attention. The video was paused between each vignette to allow time for participants to record their guesses. Aggregating over sessions, each vignette was viewed by 10–22 subjects (mean = 14.5, SD = 1.39).

Experimental Participants. Participants were undergraduate students (145 female, 73 male) enrolled at the University of Pennsylvania ($n = 159$) or La Salle University ($n = 59$) in Philadelphia. Guessing accuracy did not vary between testing locations, so participant data were pooled for all analyses [$t(216) = 0.025$, $P = 0.98$]. We chose to test participants at these two locations to increase the diversity of our participant pool. Because the families in the vignettes have a wide range of ethnicity and SES, we wanted our participant pool to have similar diversity. We used maternal education to measure the SES of our participants because many students reported that they did not know their parents' income level. Maternal education ranged from 10 y (some high school) to 20 y (advanced degree) and was higher for University of Pennsylvania students (mean = 16.91, SD = 3.02) than for La Salle students (mean = 15.59, SD = 2.75). Maternal education did not predict participants' guessing accuracy (Pearson correlation, $r = -0.004$, $P = 0.955$). All participants were native English speakers. Participants received course credit or payment for participating.

Analysis: Coding Participant Responses. Participant guesses were scored as correct if they were identical to the target word. Abbreviations and plurals were also counted as correct (e.g., phone or phones for telephone), but words that altered the meaning of the root word were not. Thus, "puppy" was not counted as a correct guess for "dog," but "dogs" was considered correct. Responses were coded as incorrect if they were either more specific than the target word (e.g., "finger" instead of "hand") or more general than the target word (e.g., "toy" instead of "bear"). These criteria were the same as those reported in ref. 7.

Participants' responses were used to calculate the guessing accuracy (proportion of correct guesses) for each video clip. The guessing accuracies of a parent's 10 target videos were then averaged to create an average HSP accuracy for that parent. This average HSP accuracy was used as the measure of quality of input for each family. Once data collection was complete, we discovered that 17 of our 500 vignettes did not accurately indicate the presence of target words via a beep, because of experimenter error. These videos were spread across 12 families and HSP accuracy for those videos was excluded from analysis.

We used linear regressions to examine the relationships between (*i*) the quantity of parent input, (*ii*) the quality of parent input, (*iii*) child vocabulary at 54 mo, and (*iv*) family SES. As described earlier, quantity of parent input was the average number of words per minute a parent directed toward her child at the 14- and 18-mo observation sessions. Quality of parent input was the average HSP accuracy for that parent derived from the experimental paradigm. Child vocabulary at 54 mo was the standardized PPVT score taken at that age. Family SES was a composite of primary caregiver education and family income. Linear regressions were used to analyze the ability of input quality and quantity to predict child vocabulary at 54 mo (Table S2).

Parent Input for Unknown Words. To ensure that HSP accuracy was measuring the quality of parent input in word-learning situations and not children's knowledge of target words, we restricted analysis to vignettes in which the children did not yet know the target words. Our measure of child knowledge of words during vignettes was based partially on children's production during observation sessions and partially on parents' completion of the MacArthur-Bates CDI. Parents were given the CDI at each observation session, but not all parents completed it at each session. To ensure that some children were not given greater opportunity than others to demonstrate their knowledge of words, we restricted analysis to families that completed the CDI at 18 mo ($n = 45$). All but one of our target words were items on the CDI ("step" was not but was a vignette for only one family).

Removing vignettes in which the child knew the target word halved the number of vignettes in the data and left some families with very few vignettes. Some children did not know any of the target words when the vignettes were recorded; others knew nearly all of them (mean number of known words = 3.66 ± 1.97). To ensure that our measure of parent quality was guessing accuracy averaged across several vignettes, we excluded three families who had only one or two videos of unknown target words. This process left us with a sample of 42 parents who had been given the CDI at 18 mo and had at least three vignettes in which their children did not know the target word. Linear regressions were used to analyze input quality (HSP for unknown words) and input quantity (words per minute) in relation to child vocabulary at 54 mo (Table S3).

Description of Known and Unknown Words. Reducing the dataset to 42 families reduced the list of target word types from 40 to 32. Of those 32 words, 25 were known (i.e., had been produced) by at least one child before the time the vignette for that child had been recorded. Overall, children in the 42 families did not know the target words in 69% (±19%) of their family's vignettes (range 30–100%). The percentage of vignettes in which the child had not yet produced the word was normally distributed across families (Shapiro–Wilk, $P = 0.118$). Only three of the target words were known by more than 50% of the children (ball, dog, book), after excluding idiosyncratic target words (i.e., those that were included in only one family's set of vignettes). Table S4 shows the number of vignettes included in the known/unknown word analysis and the number of vignettes in which the target words were known at each age (14 and 18 mo).

Our analysis of child word knowledge focused on production of the target words. Undoubtedly, the children in our videos knew more of the target words than they produced, but we did not measure vocabulary comprehension during the observation sessions at 14 and 18 mo. Parents were given the MacArthur-Bates CDI Words and Gestures at 14 mo and were asked to report the words that their children understood as well as those that they produced. Only 17 parents completed and returned the CDI Words and Gestures form, and so we could not include comprehension in our measure of child word knowledge.

One recent eye-tracking study from Bergelson and Swingley (8) found that 6- to 9-mo-old infants reliably associated pictures of objects with spoken nouns. Infant comprehension was measured using a "looking-while-listening" procedure (9), in which infants preferentially looked at pictures when they heard their parent label the objects in those pictures. All words fell into one of two categories: food items ($n = 8$) or body parts ($n = 8$). Each target item was displayed either in a pair of pictures or in a simplified scene, like objects on a tabletop (8). Ten of the 32 target words in our known/unknown vignette corpus overlapped with the words tested in this eye-tracking study (Table S4). These 10 words were not significantly more likely to be produced by the children at the time of the vignettes than the 22 words not tested in the eye-tracking comprehension study [$t(30) = 3.81$, $P = 0.71$]. Thus, we conclude that children weren't more familiar with these 10 words than with the other target words. We acknowledge that the 14- and 18-mo-olds in our vignettes were familiar with and probably understood many of the target words in our vignettes. However, they had not yet begun to produce the words. Moreover, before running our HSP study we tried to identify and exclude the use of any videos in which the child's overt actions in response to the word suggested they knew the word; if the parent had said "Where's the ball?" and the child then pulled a ball out of a bag, we would have not used such a video in the HSP study. Moreover, importantly, average HSP for unknown-word vignettes predicts vocabulary outcomes at 54 mo, but average HSP for known-word vignettes does not. This difference suggests that parents are providing richer cues to meaning before children begin producing words. It remains to be seen whether the quality of parent input shows similar changes before and after children begin to understand parent's words.

Vocabulary at Other Ages. As noted earlier, the PPVT was administered several times during the original longitudinal study (at 30, 42, and 54 mo, kindergarten, and second grade). We compared parent input quality for unknown words to child PPVT scores at all ages. Of the 56 children in our sample, 28 completed the test at the 30-mo visit, 49 at the 42-mo visit, 50 at the 54-mo visit, 49 at the kindergarten visit, and 46 at the second grade visit. Quality of parent input did not predict child PPVT at 30 mo (linear regression, $r^2 = 0.01$, $P = 0.638$) or at 42 mo ($r^2 = 0.02$, $P = 0.301$); it marginally predicted PPVT in kindergarten ($r^2 = 0.07$, $P = 0.060$) and significantly predicted PPVT at 54 mo ($r^2 = 0.12$, $P = 0.014$) and in second grade ($r^2 = 0.09$, $P = 0.042$). Quantity of parent input did not predict PPVT at 30 mo ($r^2 = 0.09$, $P = 0.126$), marginally predicted PPVT at 42 mo ($r^2 = 0.08$, $P = 0.056$), and significantly predicted PPVT in kindergarten ($r^2 = 0.09$, $P = 0.037$), at 54 mo ($r^2 = 0.13$, $P = 0.011$) and in second grade ($r^2 = 0.28$, $P < 0.001$).

Because differences in child vocabulary become more pronounced over time, early PPVT scores may not yet reflect growing individual differences in vocabulary (fewer children also contributed to the PPVT scores at 30 mo than at the other observation periods). On the other hand, once children enter school, they are exposed to a wide variety of new input that will vary between schools. Assessing child vocabulary at 54 mo right before school entry maximizes the impact that exposure to parent input can have on the measure and minimizes variation because of schooling.

Coding Behavioral Variables. Two trained coders hand-coded all vignettes using the coding software ELAN (ELAN is designed by the Max Planck Institute for Psycholinguistics, The Language Archive, The Netherlands. It is available online at http://tla.mpi.nl/tools/tla-tools/elan. A detailed description can be found in ref. 10). The coders knew the target word the parent said in each video, but the videos were muted except for the addition of beeps when the target word occurred. [Coders were informed of the target word so that they could mark presence of and attention to the target referent. We were concerned that this prior knowledge might introduce a bias in coding, so we ran a brief experiment where participants coded child attention in a subset of 20 vignettes, either knowing ($n = 10$) or not knowing ($n = 10$) the target word. Importantly, the κ-scores for coders in the two different conditions were not significantly different from each other. Agreements on parent attention between this group of coders and our coders, as assessed by κ-statistics, ranged from 0.77 to 0.87 (mean = 0.83, SD = 0.03), corresponding to "good" or "very good" agreement (11).] The coders marked the onset and offset of the following variables: (*i*) Presence of target referent, coded if and only if the target referent (e.g., a shoe) could be easily identified on screen. (*ii*) Parent attention to target referent and other objects, coded mainly by observable eye gaze, body, and head posture and, in cases where these were not readily determinable, physical interaction with a referent; if it could be determined that the parent was attending to the referent off screen, this too counted as attention to the referent. (*iii*) Parent gesture toward or manipulation of target, coded if the parent gestured at an object (e.g., pointing or showing), held the object, or otherwise manipulated the object. The original codes were transformed into binary scores for every second of the entire 40 s of each video (1 = presence; 0 = absence). Interrater reliability was assessed by having both coders record the presence and absence of the cues on 112 vignettes. The κ agreement on target presence was "near perfect," and agreements on other cues were "good" to "very good" (11).

1. Rowe ML, Goldin-Meadow S (2009) Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* 323(5916):951–953.
2. Rowe ML, Levine SC, Fisher JA, Goldin-Meadow S (2009) Does linguistic input play the same role in language learning for children with and without early brain injury? *Dev Psychol* 45(1):90–102.
3. Dunn LM, Dunn LM (1997) *Peabody Picture Vocabulary Test* (American Guidance Service, Circle Pines, MN), 3rd Ed.
4. Fenson L, et al. (1993) *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (Brookes, Baltimore, MD).
5. Goldin-Meadow S, Seligman MEP, Gelman R (1976) Language in the two-year-old. *Cognition* 4(2):189–202.
6. Gillette J, Gleitman H, Gleitman L, Lederer A (1999) Human simulations of vocabulary learning. *Cognition* 73(2):135–176.
7. Medina TN, Snedeker J, Trueswell JC, Gleitman LR (2011) How words can and cannot be learned by observation. *Proc Natl Acad Sci USA* 108(22):9014–9019.
8. Bergelson E, Swingley D (2012) At 6–9 months, human infants know the meanings of many common nouns. *Proc Natl Acad Sci USA* 109(9):3253–3258.
9. Fernald A, Zangl R, Portillo AL, Marchman VA (2008) Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, eds Sekerina I, Fernandez E, Clahsen H (John Benjamins, Amsterdam), pp 97–135.
10. Lausberg H, Sloetjes H (2009) Coding gestural behavior with the NEUROGES-ELAN system. *Behav Res Methods* 41(3):841–849.
11. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174.

Table S1. List of 40 target and 16 filler words used in the HSP experiment

Word type	Frequency rank	Vignette count
Target word		
Book	1	42
Ball	2	33
Dog	3	27
Car	4	30
Hand	5	44
Water	6	35
Nose	7	39
Kiss	8	37
Shoe	9	29
Mouth	10	43
Fish	11	11
Bear	12	13
Eye	13	18
Door	14	17
Bird	15	2
Hair	16	12
Cat	17	2
Horse	18	1
Foot	19	6
Duck	20	1
Head	21	6
Juice	22	5
Cookie	23	6
Milk	24	4
Pig	25	3
Block	26	2
Chair	27	2
Apple	28	1
Phone	29	1
Cup	31	1
Sock	32	2
Orange	33	1
Cow	34	1
Step	35	1
Button	37	1
Cheese	40	1
Bed	62	1
Shirt	66	1
Bowl	103	1
Bread	115	1
Filler word		
Mom	1	48
Job	2	37
Two	3	43
Bite	4	31
Dance	5	33
Blue	6	19
Time	7	18
Toy	8	10
Music	9	2
Thing	10	2
Picture	11	1
Today	12	1
Color	13	1
Outside	14	1
Piece	15	1
Floor	16	1

Words are ranked according to their frequency in all parent speech in the 90-min transcripts at 14 and 18 mo. The vignette count column indicates how many vignettes (and thus parents) used that word. Data from the 50 families included in the analysis are reported.

Table S2. Multiple linear regression using parent words per minute and average HSP accuracy to predict child PPVT score at 54 mo

Model	Variable	*r*	r^2	Unstandardized coefficients (β)	Unstandardized coefficients SE	Standardized coefficients (β)	*t*	*P*
1	Constant			99.593	5.420		18.374	0.000
	Parent words per minute	0.357	0.127	0.336	0.127	0.357	2.644	0.011
2	Constant			84.515	8.131		10.394	0.000
	Parent words per minute			0.304	0.122	0.323	2.496	0.016
	HSP accuracy	0.472	0.223	75.231	31.316	0.311	2.402	0.020

Model 1 uses parent words per minute (input quantity) to predict child PPVT score (vocabulary) at 54 mo. Model 2 uses both parent words per minute and average HSP accuracy (input quality) as predictors. The r^2 values for the models are: r^2(model 1) = 0.127 and r^2(model 2) = 0.223.

Table S3. Multiple linear regression using parent words per minute and average HSP accuracy for unknown words to predict child PPVT score at 54 mo

Model	Variable	*r*	r^2	Unstandardized coefficients (β)	Unstandardized coefficients SE	Standardized coefficients (β)	*t*	*P*
1	Constant			99.001	6.323		15.657	0.000
	Parent words per minute	0.384	0.148	0.386	0.146	0.384	2.633	0.012
2	Constant			88.732	7.902		11.229	0.000
	Parent words per minute			0.356	0.142	0.355	2.512	0.016
	HSP accuracy unknown words	0.479	0.230	61.070	29.965	0.288	2.038	0.048

Analysis is restricted to the 279 vignettes that contained target words unknown to the child. Model 1 uses parent words per minute (input quantity) to predict child PPVT score (vocabulary) at 54-mo. Model 2 uses both parent words per minute and average HSP accuracy of vignettes containing unknown words as predictors. The r^2 values for the models are: r^2(model 1) = 0.148 and r^2(model 2) = 0.230.

Table S4. Target word production at 14 and 18 mo

Target word	Frequency rank	Total vignette count	In known/unknown dataset	Vignettes 14 mo	Produced (known) 14 mo	Vignettes 18 mo	Produced (known) 18 mo	Eye-tracking comp
Book	1	42	38	16	5	22	15	
Ball	2	33	30	16	8	14	13	
Dog	3	27	22	12	4	10	8	
Car	4	30	26	17	1	9	3	
Hand	5	44	40	21	1	19	3	Yes
Water	6	35	32	20	4	12	3	
Nose	7	39	33	15	1	18	10	Yes
Kiss	8	37	31	10	0	21	9	
Shoe	9	29	24	14	3	10	7	
Mouth	10	43	36	21	0	15	5	Yes
Fish	11	11	10	6	1	4	1	
Bear	12	13	13	8	2	5	0	
Eye	13	18	14	7	2	7	3	Yes
Door	14	17	14	8	0	6	1	
Bird	15	2	1	0	0	1	0	
Hair	16	12	7	2	0	5	2	Yes
Cat	17	2	2	0	0	2	1	
Horse	18	1	0	0	0	0	0	
Foot	19	6	4	1	0	3	1	Yes
Duck	20	1	1	0	0	1	1	
Head	21	6	4	0	0	4	0	
Juice	22	5	3	3	1	0	0	Yes
Cookie	23	6	5	3	1	2	1	Yes
Milk	24	4	4	3	1	1	1	Yes
Pig	25	3	3	1	0	2	1	
Block	26	2	1	1	0	0	0	
Chair	27	2	2	1	0	1	0	
Apple	28	1	1	1	1	0	0	Yes
Phone	29	1	1	0	0	1	1	
Cup	31	1	1	1	0	0	0	
Sock	32	2	1	0	0	1	1	
Orange	33	1	1	1	0	0	0	
Cow	34	1	0	0	0	0	0	
Step	35	1	1	1	0	0	0	
Button	37	1	0	0	0	0	0	
Cheese	40	1	0	0	0	0	0	
Bed	62	1	0	0	0	0	0	
Shirt	66	1	0	0	0	0	0	
Bowl	103	1	0	0	0	0	0	
Bread	115	1	0	0	0	0	0	

The number of vignettes included in the known/unknown word analysis is listed to the right of the vignette count column. These smaller values reflect the exclusion of vignettes from eight families who either did not complete a CDI at 18 mo, or whose child did not have more than three unknown word vignettes (i.e., knew the words in almost all of the vignettes). Data from the 42 families included in the unknown word analysis are reported. The number of vignettes from the 14- and 18-mo observation sessions is given for each of the 40 target words. The "produced" columns show the number of vignettes recorded following child production of the target word (i.e., we observed production or the parent reported it). The "eye-tracking comp" column indicates those target words that were included in an eye-tracking study of word comprehension in 6- to 9-mo-old infants by Bergelson and Swingley (7). This eye-tracking study showed that infants preferentially looked toward pictures of these items when their parents said one of these words, but the design did not reveal which words individual children knew. Thus, we cannot determine the exact probability that the children in our study understood any one of these words. However, we can assume that some of the children in our study understood some of these words.