Cross-cultural differences in children's object handling at home

Marisa Casillas[1,2] & Mary Elliott[3,2]

[1] University of Chicago

[2] Max Planck Institute for Psycholinguistics

[3] University of Texas at Dallas

Abstract

Object-centered interactions (e.g., child-caregiver toy play) are thought to make significant contributions to children's early word learning. However, it is yet unknown how frequently such interactions occur in children's daily lives. We investigate how often 83 children under age four (and their interactants) handle objects during everyday life in two non-Western communities: one Mayan and one Papuan. Indeed, after infancy, children handle objects relatively frequently (15% or more of their time) and do so in bursts and lulls across the day. Importantly, cultural differences impact how frequently children handle objects across age, perhaps due to differences between these communities in infant carrying practices, young children's daily activities, and the objects available to children. In contrast, we find little evidence to support the idea that object handling by children's interactants might similarly drive early word learning. We discuss the implications of these findings for theoretical and computational models of word learning.

Keywords: culture, word learning, Mayan, Papuan, daylong recording, egocentric images

Word count: X

Cross-cultural differences in children's object handling at home

Statement of relevance

Before children go to school, they learn many words at home. Object-centric interactions may propel this early learning: communication is facilitated when the child and an interactant jointly focus on an object (e.g., a toy). However, it is unknown how often children actually engage in such interactions at home, nor whether these interactions are similarly present across diverse cultural contexts. We investigate how often young children and their interactants handle objects during daily life in two Indigenous communities: one Mayan and one Papuan. We find that, while children themselves often handle objects, their interactants rarely handle objects immediately relevant to them. Children also showed different object handling patterns between communities, likely due to differences in infant carrying style and child daily activities. Therefore, focusing on child-led (not parent-led) object interactions and attending to how cultural factors shape everyday activities will be key to illuminating how children learn words at home.

Introduction

As children gain the ability to pick things up, sit on their own, and move around independently, they also experience immense changes in their social and object-centered interactions (Adolph, Karasik, & Tamis-LeMonda, 2010; Franchak, Kretch, Soska, & Adolph, 2011; Gaskins, 2000; Kretch, Franchak, & Adolph, 2014; Sanchez, Long, Kraus, & Frank, 2018). Object-centric interaction with others, particularly coordinated child-caregiver attention via object handling, has been proposed as a potentially powerful source of information for early word learning (e.g., Yu & Smith, 2013). Egocentric recordings of interaction show decreasing views of faces and increasing views of hands over the first two years of life (Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2017; but see Long, Kachergis, Agrawal, & Frank, 2020). This focus on active hands can lead to episodes of shared attention in which communication, and thereby word learning, is facilitated (Yu &

Smith, 2013). And while both children and their caregivers can initiate these interactional episodes, children's own holding experiences throw objects into sharp perceptual relief, thereby increasing the potential for learning about that object's label, functional affordances, and more (e.g., Amatuni et al., 2021; Elmlinger, Suanda, Smith, & Yu, 2019; Slone et al., 2018; Soska, Adolph, & Johnson, 2010; see also Clerkin, Hart, Rehg, Yu, & Smith, 2017; Lockman & Kahrs, 2017; Long, Kachergis, Bhatt, & Frank, 2021).

In order to establish such episodes as plausible and universal drivers of early word learning, we must determine how often they occur during children's daily lives and how their distribution changes across cultural contexts and child age. If object handling interactions play a critical role in word learning, we would expect them to (a) occur frequently, to enable the accumulation of evidence across multiple timepoints; and (b) in massed (i.e., bursty) distribution, at least for younger children, who may struggle to learn labels when the information is spread over longer periods (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Long, Kachergis, Agrawal, & Frank, 2020; Vlach & Johnson, 2013; Yurovsky & Frank, 2015; Yurovsky, Smith, & Yu, 2013; Zhang, Yurovsky, & Yu, 2021). Finally, (c) we would need to see that these distributional properties are robustly maintained across diverse cultural contexts.

We do not yet know how often such opportunities for hand-led intersubjective attention arise during children's day-to-day lives around the world. In a free-play setting in the lab, US infants and their caregivers handle objects over 90% of the time (Yu & Smith, 2013). However, at-home recordings suggest much lower handling rates, with the appearance of hands in children's view topping out around 30% (Fausey, Jayaraman, & Smith, 2016; Long, Kachergis, Agrawal, & Frank, 2020). While at-home recordings effectively capture a variety of activity contexts, they have tended to be short and parent-selected in past work, so are not representative of children's whole waking days. Further, while prior work has focused on North American children, cross-cultural differences in how children are carried, where they

are placed, what kinds of objects typically surround them, and what activities they engage in guarantee wide variability in object handling (Adolph, Karasik, & Tamis-LeMonda, 2010; Gaskins, 2000). For example, children who spend much of their first year carried or tightly bound (e.g., for safety or warmth, Hayashi, 1992; Ishak, Tamis-LeMonda, & Adolph, 2007; LeVine & Lloyd, 1966; Mei, 1994) cannot easily reach out to pick up nearby objects.

The present study examines natural patterns of child and interactant object handling at home in two unrelated, non-Western populations. We analyzed the frequency and distribution of object handling in more than 113000 child-perspective photos taken during 83 daylong, at-home recordings of children's waking days at home in two communities: one in which children are typically carried on their mother's back for most waking hours during the first year of life (Tseltal; Mayan) and one in which children are typically carried in caregivers' arms during the first year (Rossel; Papuan; Figure 1). Our results suggest that children—but not their interactants—indeed handle objects fairly frequently from toddlerhood onward, with handling episodes distributed in bursts and lulls across the day. That said, object handling patterns across age varied by population, pointing to potentially important effects of cultural context in explaining how object-centric interaction might drive human language development. In what follows we describe each population and explain the methods for data collection, annotation, and analysis before detailing the results. We discuss the implications of the present findings for both theoretical and computational models of early word learning.

Method

Participating communities

We first describe the typical carrying practices, physical and social environments, and patterns of child-directed speech observed in the two communities of study. Neither author is a member of the communities described. Our description is informed by experience talking to and observing families, plus consultation with two researchers who have worked there for several decades (P. Brown and S. C. Levinson). See Supplementary Materials for example

Figure 1. Typical infant holding positions in the studied Tseltal and Rossel communities.

images of scenes typical to each context.

The Tseltal (Mayan) participants live in a rural swidden horticulturalist community situated within the Chiapas highlands of Southern Mexico. For most of the first year of life, children are carried in a sling on their mother's back during her waking hours and are rarely put on the ground before they walk (Brown, 2011, 2014). When sleeping they are wrapped up completely in the sling, and when waking they are hitched into a sitting position with the head and torso free (Figure 1). We have also observed infants sometimes being carried in the arms, or held on the waist or in a lap. When children *are* placed on the ground, it is usually on a woven mat or blanket, or occasionally in a cardboard box, and always under the surveillance of a nearby caregiver. Infants are occasionally put to sleep in a hammock during the daytime. For these reasons, it is rare to see Tseltal infants crawling. Mothers report that, at some point, infants begin making limb movements while tied up in the sling, indicating their readiness to begin standing and, soon after, taking their first steps. At this point, caregivers may provide standing and walking support with their bodies and household objects until the child is walking on their own. This Tseltal community is situated on a

mountainside such that, when children and their caregivers leave the home (e.g., to visit a relative), they must typically traverse steep dirt paths for part of the way, occasionally walking on a paved road. Thus, while nearly all children between 12 and 18 months walk, they may still be carried for significant parts of the day, especially over longer distances or challenging terrain.

The Rossel (Papuan) participants in our study live in a collection of swidden horticulturalist communities on the northeastern region of Rossel Island, which is the farthest outlying atoll of the Louisiade Archipelago, off the coast of mainland Papua New Guinea. Like Tseltal children, Rossel children are carried for much of their first year. However, they are typically carried in the arms: That is, on the waist, against the chest, and over the shoulder (Figure 1, Brown, 2011; Brown & Casillas, accepted). Children, even young infants, are cared for by a wide network of nearby family members and neighbors (male and female, adult and child), any of whom might carry or hold the infant for sustained periods. Rossel infants, like Tseltal infants, are rarely put directly on the ground. Instead they are more often found laying, sitting, scooting, and crawling on the raised verandas that are constructed under or attached to residences. Similarly, infants are sometimes placed in hammocks for daytime sleep. Once children show signs of interest in walking, caregivers might set out a line of small posts in the ground that the infant can grasp and walk along under a caregiver's supervision. In brief, while infants are rarely placed on the ground in both communities, Tseltal children's movements and ability to grasp nearby objects is more restricted early in development.

Cross-cultural differences in carrying practices less strongly influence children's object handling once they begin to walk, but the social landscape and available objects create persistent differences between communities. While children in both contexts spend significant time interacting with other children (e.g., older siblings and cousins/neighbors), this pattern is especially prominent in the Rossel community, where children join large,

independent child playgroups shortly after they start to walk (Brown, 2011, 2014; Brown & Casillas, accepted). These large playgroups sometimes engage in stationary, object-centric, verbally interactive activities (e.g., pretend household play, cracking and eating foraged nuts), but more often facilitate mobile activities in which few objects are relevant and verbal activity is repetitive or routinized (e.g., diving games in the river, chasing games similar to 'tag'). Tseltal children tend to participate in smaller child social groups and tend to move within a somewhat more restricted area around their household grounds. Spot observations of Yucatec Mayan children suggest that play, particularly manipulative play with objects and substances, increases in frequency with age, to 40% of the child's time by ages 3–5 (Gaskins, 2000, work activities, which may involve object manipulation, also increase in this period). While both communities are remote, the Tseltal community is far more commercially connected to the Western, industrial world—Rossel Island sees only infrequent and irregular boat contact. For this reason, objects that are designed specifically for children's interest or manual manipulation (e.g., toys, crayons, etc.) are even less frequently seen in the Rossel context than the Tseltal one.

Recording methods

The present data come from a subset of daylong photo-linked audio recordings collected in 2015 (Tseltal) and 2016 (Rossel) of 55+ children under age 5;0 in each site. Children were outfitted with an elastic vest that carried two devices: A lightweight Olympus audio recorder (WS-832 or WS-853) and a wearable camera (Narrative Clip 1) with a miniature fisheye lens (Photojojo Super Fisheye; Figure 2). Some infants (typically those 0;6 and younger) could not wear both devices at once, and so instead wore an infant bodysuit ("onesie") with the audio recorder while their caregiver wore an adult-sized vest with the camera.

The camera captured images at a fixed interval over the course of the recording day at home, which typically lasted around 9 (Tseltal) or 8 (Rossel) hours. Timestamped images were captured every 30 seconds in the Tseltal data and, following a camera firmware update

in late 2015, every 15 seconds in the Rossel data. Participants were able to cover the camera lens at their discretion using a piece of cloth attached to the underside of the camera case (see Casillas, Brown, & Levinson, 2020 for details).

Figure 2. Narrative clip camera with attached fisheye lens.

Information about the child's developmental, linguistic, domestic, and demographic profile was collected via interviews between the child's caregivers, the first author, and a research assistant who lives in the community. Dates of birth were collected both verbally and from any available medical documentation. When dates were inconsistent across information sources, we triangulated them using any additional information we could gather (e.g., birth date relative to another child).

We here analyze photo data from 40 Tseltal and 43 Rossel daylong recordings (from 38 and 42 children, respectively). We focused on children between ages 0;0 and 4;0, which was the target age range during data collection, and hence the most densely sampled. Across sites, the children are balanced as well as possible in terms of age and sex though imbalances remain. Our samples incorporate a majority of the local population in the target age range at the time of data collection.

Annotation

We annotated photos with IMCO (github.com/marisacasillas/ImCo/), an open-source program that allows researchers to efficiently annotate photos using keyboard inputs that are mapped to predefined categories. Each photo dataset typically took a trained research assistant fewer than 10 seconds to annotate. For each photo we annotate: The number of visible adults (i.e., post-pubescents) and children, whether any visible child was crying or breastfeeding, whether the target child was handling an object, whether one or more of the target child's interactants was handling an object directly relevant to the target child (e.g., while feeding them), and whether the photo was unusable or skipped (e.g., due to overexposure). We only annotated photos between the time the researcher left and the time she returned because caregivers tended to restrict the target child's movements during this period.

The present study examines object handling, for which there are four categories: 'C' = the target child is handling an object (e.g. while playing with a toy); 'I' = one or more of the target child's interactants are handling an object relevant to the child (e.g. while washing or feeding the child); 'B' = the target child *and* one or more of the target child's interactants are handling an object (e.g. while playing with a toy together); 'N' = there is no object handling visible.

A handled object was defined as something (e.g. a toy, piece of food, or rock) that the target child was holding or manipulating. Large or immovable objects were considered handled if the child was actively engaging with them (e.g., a branch while climbing a tree, but not a table on which a hand was resting). People were not counted as handled objects (e.g. a mother holding her baby or a child holding a breast). Objects near, but not directly in contact with, a child's hands were coded as handled when justifiable (e.g., a hand reaching toward a ball rolling toward it). The chest-worn cameras were also considered held when handled by the participants.#apostrophe

Reliability and data preparation

We analyze 151827 annotations of 113668 photos (41064 from Tseltal recordings and 72604 from Rossel recordings) by three annotators. A substantial proportion of photos from each site were annotated by at least two annotators (Tseltal: 0.44; Rossel: 0.24). One annotator contributed the majority of annotations for both sites (Tseltal: 29528; Rossel: 50331 unique photos), while the annotators primarily on one site or the other (Tseltal: 9557 and Rossel: 39779 vs. Tseltal: 22632 and Rossel: 0).

We analyze those photos deemed 'usable' by at least one annotator (87949 of the photos; 77.40% of the full annotated set). The remaining 25719 photos were blocked by the provided privacy cover, had too bright/dark lighting to annotate, were taken while the researcher was present, or were otherwise uncodeable. On average, this left us with 1,058.95 photos per recording, but with wide variability between recordings (median: 1044; range: 13–1737).

Annotator agreement was high (85.90%), with comparable scores for Tseltal and Rossel recordings (Tseltal: 87.80%; Rossel: 83.10%). The primary source of disagreement came from whether the target child was handling an object or not ('C' vs. 'N' disagreements; see Supplementary Materials). Photos with no visible object handling were by far the most common outcome, so given cases of disagreement, the resulting unweighted Cohen's kappa score suggests moderate overall agreement (kappa = 0.59).

We derived a single annotation value for each photo as shown in Table 1. We then derived a single burstiness estimate for child ("C" or "B") and interactant ("I" or "B") object handling for each recording, based on Goh and Barabási (2008)'s B parameter. Burstiness (B) is calculated as $B = (\sigma_\tau - \mu_\tau)/(\sigma_\tau + \mu_\tau)$ where τ is the distribution of inter-event intervals (IEIs[1]). A B score of -1 indicates a completely periodic distribution; 1 a

[1] An "event" here is a photo featuring object handling. Because of the coarse nature of the photo data (i.e.,

Table 1

A single annotation value was derived for each photo from the combination of annotations provided by coders as described below.

Combination	Result	# Photos
B only	B	566
C + C/B	C	14011
I + I/B/C	I	982
N + anything	N	72390

maximally bursty distribution; and 0 a random distribution. For reference, multimodal event distributions in adult matcher-director games have been found to typically fall around $B = 0.15$–0.2. (Abney, Dale, Louwerse, & Kello, 2018).

Results

All analyses and figures were generated in R (Aust & Barth, 2020; R Core Team, 2020; Wickham et al., 2019). The anonymized data and analysis code are available at github.com/marisacasillas/daylong-photos. Full tabular regression outputs can be found in the Supplementary Materials.

Frequency and distribution of object handling.

Children.

Across both daylong photo datasets, children 4;0 and under handled objects in an average of 15.80% of photos (median = 12.90%, range = 0%–56.73%). Prevalence of object

samples every ~15 or ~30 seconds) we cannot precisely say when handling events started and stopped. IEI is therefore the time in seconds between two consecutive photos featuring object handling of the desired type (here "C"/"B" or "I"/"B" for the child and interactant analysis, respectively).

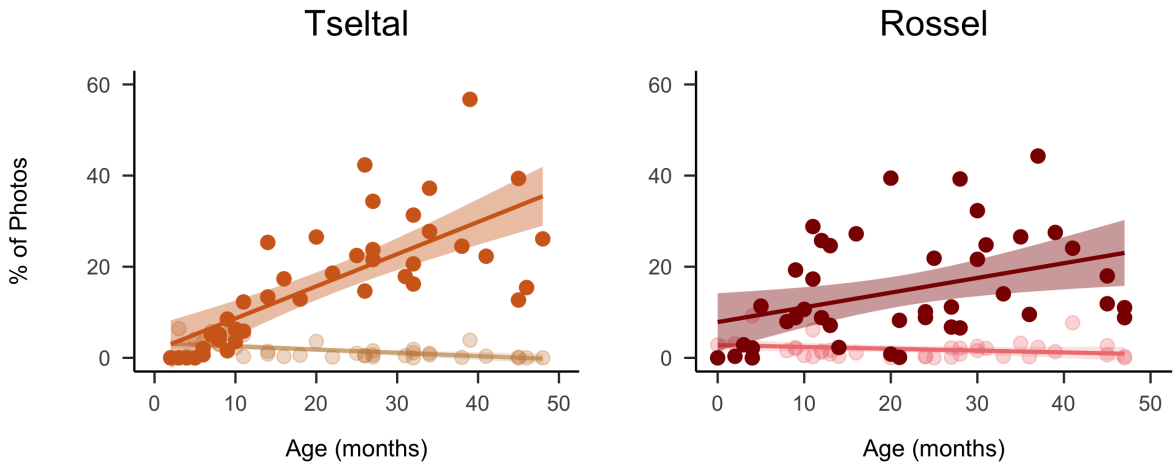

Figure 3. Proportion of photos in which children (dark) and their interactants (light) are seen handling an object during the Tseltal (left) and Rossel (right) daylong recordings.

handling was overall similar between sites, with an average of 16.91% of photos showing object handling for the Tseltal dataset (median = 15.82%, range = 0%–56.73%) and 14.77% of photos for the Rossel dataset (median = 11.02%, range = 0%–44.29%). However, inspection of object handling by age across sites reveals a pattern whereby Tseltal infants handle objects less frequently than Rossel children early on, but then do so more frequently than Rossel children later (Figure 3).

A linear regression modeling the effects of age, cultural group, and their interaction on the prevalence of child object handling (i.e., the proportion of photos in which the child was handling an object; N = 83; AIC = -136.73)[2] revealed impacts of both age and cultural group. Specifically, children handled objects more frequently with age overall ($b = 0.003$, 95% CI [0.002, 0.004], $t(79) = 2.802$, $p = 0.006$), and children in the Tseltal data based showed a larger increase in object handling with age compared to Rossel children ($b = 0.004$, 95% CI [0.002, 0.005], $t(79) = 2.322$, $p = 0.023$).

[2] glm(Proportion of photos showing object handling ~ Child age (months; numeric) * Site (Tseltal/Rossel; factorial))

Children's object handling was bursty in all 77 recordings with two or more instances of child object handling (Figure 4); B_{mean} = 0.45, B_{sd} = 0.17, B_{range} = -0.05–0.81). A linear regression[3] (N = 77; AIC = -48.63) revealed no effects of age or cultural context on burstiness.

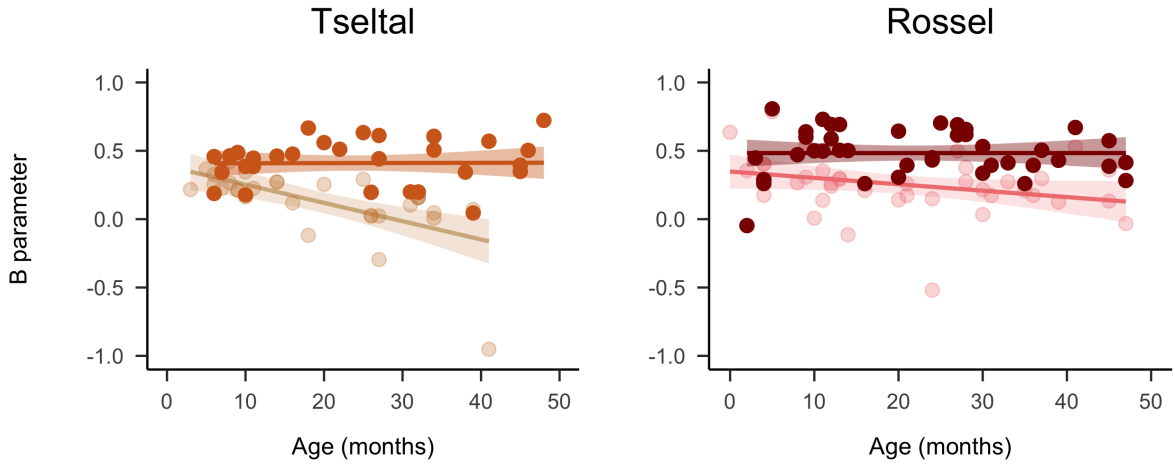

Figure 4. Burstiness values for child (dark) and their interactants' (light) object handling during the Tseltal (left) and Rossel (right) daylong recordings.

Interactants.

Overall, children's interactants handled child-relevant objects in an average of 1.82% of photos (median = 1.09%, range = 0%–11.01%). Prevalence of interactant object handling was also similar between sites, with an average of 1.70% of photos for the Tseltal dataset (median = 0.92%, range = 0%–6.41%) and 1.94% for the Rossel dataset (median = 1.36%, range = 0%–11.01%). Interactant object handling, unlike child object handling, appears descriptively similar across age between sites (Figure 3). Using the same model structure, we found no evidence to support impacts of age, cultural group, or their interaction on interactant object handling (N = 83; AIC = -405.52; all $|t| < 0.54$).

Interactants' object handling was bursty in 63 of the 69 recordings with two or more instances of interactant object handling (0.91; Figure 4; B_{mean} = 0.20, B_{sd} = 0.24, B_{range} =

[3] glm(B ~ Child age (months; numeric) * Site (Tseltal/Rossel; factorial))

-0.95–0.79). A linear regression (N = 69; AIC = -15.22) revealed a significant age-by-site interaction whereby burstiness decreased more across age for Tseltal interactants than Rossel ones (b = -0.009, 95% CI [-0.013, -0.005], $t(65)$ = -2.113, p = 0.038). There was no evidence for further effects of age or cultural context.

Discussion

The present data suggest that object handling episodes have the greatest potential for word learning after age one, though this may vary across cultural contexts; the data do *not* support interactant object handling as critical for early word learning. Specifically, children's object handling increased with age, reaching 15% of the time or higher for children over age 1;0, but showed differing developmental trajectories between sites—Tseltal children initially handled objects less often than Rossel children did, but then later handled objects more often. Interactants' child-relevant object handling was very rare—around 1–2% of the time—and was not impacted by age or cultural context. With little exception, handling events occurred in bursty distribution across the day by both children and interactants, and across age and cultural contexts.

Cross-cultural differences

The difference in developmental trajectories between these two sites is likely driven by multiple factors, including carrying practices and the social and economic organization of daily life. Carrying practices may drive the differences that appear before children begin to walk; Tseltal infants under 1;0, who are carried in a sling, seldom handled objects, while Rossel infants, who are typically carried in the arms, steadily increased object handling over the first year. Social organization impacts the number and composition of interactants present, as well as the types of activities children engage in; in our dataset, children on Rossel Island typically spent long stretches of the day with large, independent play groups that frequently engaged in non-object-centric activities (e.g., playing tag, diving in the river) whereas Tseltal children tended to spend more of their day doing domestic themed,

object-centric activities (e.g., (playing at) preparing foods, drawing and/or toy play) in smaller groups, and stayed nearer to the immediate area around the family home. More market integration in the Tseltal community with its surrounding industrialized economies also translates to a much higher preponderance of toys and other small manmade objects compared to Rossel Island (e.g., cups, pens, plastic bottles, etc.), though we note these objects are still less prevalent than they are in, e.g., the EuroAmerican homes where much of the prior work on object-centric interaction has taken place.

How much is enough?

Establishing naturalistic home rates of object handling in interaction will helpfully constrain the input assumptions made by theoretical and computational models of word learning in the first few years of life. In-lab investigations of US object-centric interactions show objects being held by children or their interactants over 90% of the time (Yu & Smith, 2013), which is not sustainable during everyday life at home, including housework activities, adults socializing, daytime sleep periods, and so on, all of which are captured in home daylong recordings. The present results place the likely frequency of object handling closer to indirect estimates based on the general prevalence of hands during short, at-home recordings—a maximum of around 25–30% of the time for children over 1;0 (Fausey, Jayaraman, & Smith, 2016; Long, Kachergis, Agrawal, & Frank, 2020).

Importantly, we do not know how much object handling is "enough" to support early word learning. If this type of learning episode is very potent or frequently accompanied by actual object labels, even low rates of object handling may result in robust word learning. Given the relatively low rates of child-directed talk during at-home recordings in these and Western communities (Bunce et al., 2020; Casillas, Brown, & Levinson, 2020, 2021) we anticipate that object handling episodes that are actually accompanied by talk about the object are very infrequent. Object handling activity is, at least, bursty; repeated or sustained handling may provide a boost to label learning, compensating for the low overall frequency

of these events.

Lack of interactant object handling. Interactant object handling was rare in the present data, suggesting that in these communities it plays, at most, a minor role in early word learning. Our estimates stand in stark contrast to those from in-lab object-centric child-caregiver interaction in the US, in which adult caregivers and young one-year-olds handle objects with approximately equal prevalence (each 25% of the time on their own and 43% of the time jointly, Yu & Smith, 2013). Informally, we note that interactant-held objects were typically limited to a few prototypical items associated with basic care, e.g., food, clothing, and daily hygiene—not toys, books, or other items that would more often be found in recordings with middle-class Western families. If present data are closer to children's true at-home experiences, in Western homes or elsewhere, theories of word learning via hand-led intersubjective episodes should focus on cases when the child is doing the object handling, and not their interactant(s).

Looking ahead

The present work provides preliminary benchmarks for computational modeling and future comparative work on word learning during multimodal interaction. TThe findings also form a basis for further theory development around how object-handling events and early word learning are influenced by the child's cultural and socioeconomic milieu in concert with their motor development. Beyond word learning, naturalistic object handling patterns have implications for children's development of visual, tactile, and event-based representations of the world (e.g., the canonical form and function of familiar objects and the expected form and function of novel objects). Continued work with cross-cultural collections of highly naturalistic egocentric data will be key to understanding how children come to learn about the world, and how their learning process is continually reshaped by features of both their home environment and their own developmental gains.

Contributions

MC developed the study concept, collected the data, and contributed funding. ME annotated the data and trained other annotators. Both contributed to the writing and approved the final version of this manuscript.

Acknowledgements

We are indebted to Rebeca Guzmán López, Humbertina Gómez Pérez, Taakêmê Ńamono, Y:aaw:aa Pikuwa, and the participating families and community leaders in each population sampled. We also thank the PNG National Research Institute, the Administration of Milne Bay Province, and the Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS) Sureste. Our work builds on that of Penelope Brown and Stephen C. Levinson who were invaluable consultants in conducting this study. We are grateful to Shawn C. Tice for his technical contributions to the photo-annotation tool, and to Ine Alvarez van Tussenbroek and Cielke Hendriks for their help with annotation. This work is supported by a NWO Veni Innovational Scheme grant (275-89-033) to MC and by fieldwork funding from the Max Planck Institute for Psycholinguistics.

References

Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and non-verbal communication. *Cognitive Science*, *42*(4), 1297–1316.

Adolph, K. E., Karasik, L. B., & Tamis-LeMonda, C. S. (2010). Motor skill. In M. H. Bornstein (Ed.), *Handbook of cultural developmental science* (pp. 61–88). Psychology Press: New York, NY.

Amatuni, A., Schroer, S., Zhang, Y., Peters, R., Reza, Md. A., Crandall, D., & Yu, C. (2021). Characterizing the object categories two children see and interact with in a dense dataset of naturalistic visual experience. In *Proceedings of the 43rd annual conference of the cognitive science society* (pp. 265–271).

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & and Bambi B Schieffelin (Eds.), *Handbook of Language Socialization* (pp. 29–55). Malden, MA: Wiley-Blackwell.

Brown, P. (2014). The interactional context of language learning in Tzeltal. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in interaction: Studies in honor of Eve V. Clark* (pp. 51--82). Amsterdam, NL: John Benjamins.

Brown, P., & Casillas, M. (accepted). Childrearing through social interaction on Rossel Island, PNG. In A. J. Fentiman & M. Goody (Eds.), *Esther Goody revisited: Exploring the legacy of an original inter-disciplinarian* (pp. XX–XX).

New York, NY: Berghahn.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., ... Casillas, M. (2020). *A cross-cultural examination of young children's everyday language experiences.*

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a papuan community. *Journal of Child Language*, *48*(4), 792–814.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055.

Elmlinger, S. L., Suanda, S. H., Smith, L. B., & Yu, C. (2019). Toddlers' hands organize parent-toddler attention across different social contexts. In *2019 joint IEEE 9th international conference on development and learning and epigenetic robotics (ICDL-EpiRob)* (pp. 296–301). IEEE.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750.

Gaskins, S. (2000). Children's daily activities in a Mayan village: A culturally grounded description. *Cross-Cultural Research*, *34*(4), 375–389.

Goh, K.-I., & Barabási, A.-L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, *81*(4), 48002.

Hayashi, K. (1992). The influence of clothes and bedclothes on infants' gross motor development. *Developmental Medicine & Child Neurology*, *34*(6), 557–558.

Ishak, S., Tamis-LeMonda, C. S., & Adolph, K. E. (2007). Ensuring safety and providing challenge: Mothers' and fathers' expectations and choices about infant locomotion. *Parenting: Science and Practice*, *7*(1), 57–68.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology*, *53*(1), 38.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518.

LeVine, R. A., & Lloyd, B. B. (1966). *Nyansongo: A Gusii community in Kenya*. Wiley.

Lockman, J. J., & Kahrs, B. A. (2017). New insights into the development of human tool use. *Current Directions in Psychological Science*, *26*(4), 330–334.

Long, B., Kachergis, G., Agrawal, K., & Frank, M. C. (2020). *Detecting social information in a dense database of infants' natural visual experience*.

Long, B., Kachergis, G., Bhatt, N., & Frank, M. C. (2021). Characterizing the object categories two children see and interact with in a dense dataset of naturalistic visual experience. In *Proceedings of the 43rd annual conference of the cognitive science society* (pp. 279–285).

Mei, J. (1994). The northern chinese custom of rearing babies in sandbags: Implications for motor and intellectual development. In J. H. A. van Rossum & J. I. Laszlo (Eds.), *Motor development: Aspects of normal and delayed development*. VU Uitgeverij: Amsterdam, NL.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sanchez, A., Long, B., Kraus, A. M., & Frank, M. C. (2018). Postural developments modulate children's visual access to social information. In *Proceedings of the 40th annual conference of the cognitive science society* (pp. 2412–2417).

Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C., Franchak, J. M., Pearcy, D., ... Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *Journal of Visualized Experiments*, (141).

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*(1), 129–138. https://doi.org/10.1037/a0014618

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand

coordination. *PloS One*, *8*(11), e79659.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966.

Zhang, Y., Yurovsky, D., & Yu, C. (2021). Cross-situational learning from ambiguous egocentric input is a continuous process: Evidence using the human simulation paradigm. *Cognitive Science*, *45*(7), e13010.

Supplementary Materials: Cross-cultural differences in children's object handling at home

Marisa Casillas[1,2] & Mary Elliott[3,2]

[1] University of Chicago

[2] Max Planck Institute for Psycholinguistics

[3] University of Texas at Dallas

Supplementary Materials: Cross-cultural differences in children's object handling at home

Developmental context

Figure 1. Four images illustrating aspects of the Tseltal and Rossel environments. Left: Tseltal target infant in the mother's lap at a family morning meal in the kitchen hut. Left-center: Tseltal mother walking with a man up a main path (target child is in sling). Right-center: Rossel children using stones to crack open foraged nuts under the stairs entering a household structure. Right: Rossel child playing in the net hammock under their house with a caregiver laying nearby on the veranda.

(#fig:context.imgs)

Reliability

Confusion matrices

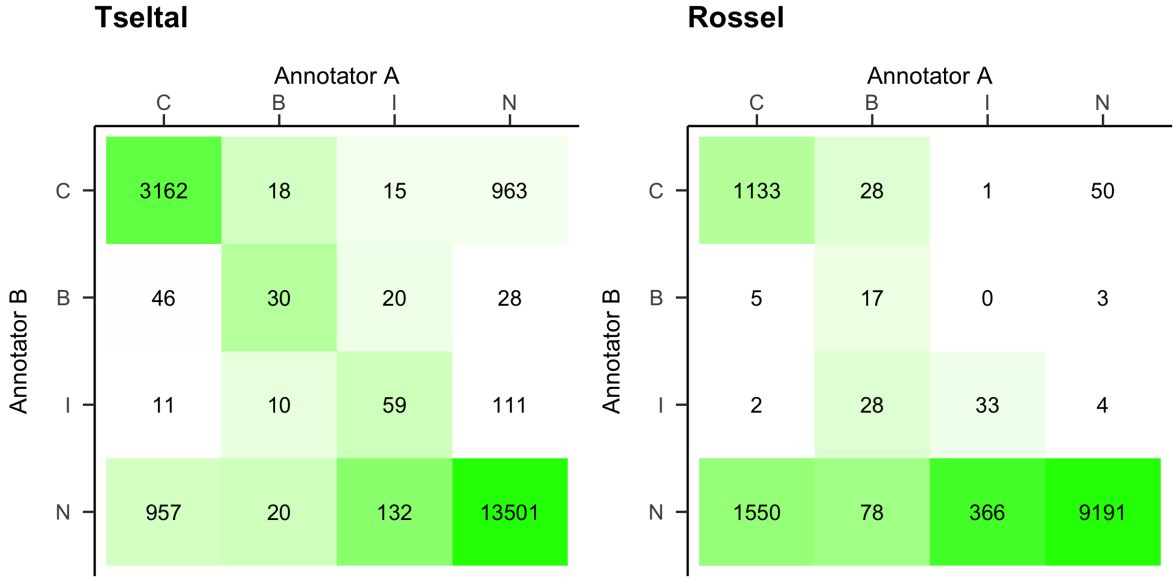

Figure 2. Confusion matrices of possible photo annotation values by site: Tseltal (left) and Rossel (right). The value in each cell indicates the number of photos with the given combination of annotations from person A and person B. The shading indicates proportion of the label category for Annotator A distributed over Annotator B's possible responses. (#fig:confmat.fig)

Table 1

Linear regression output for the analysis of annotation agreement by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.90	0.04	22.55	0.00
Age	0.00	0.00	-1.84	0.08
SiteMayan	0.02	0.05	0.44	0.66
Age:SiteMayan	0.00	0.00	0.20	0.84

Effects of age and cultural context

To test whether annotation was more reliable at different ages or in different sites, we conducted a linear regression of mean agreement value by child age (in months; numeric), site (Tseltal vs. Papuan; factorial), and their interaction (Mean_agmt ~ Age * Site; N = 29; AIC = -70.13). We found no evidence for significant impacts of age or cultural context on annotation agreement scores.

Full regression output of object handling analyses

Below are the full model outputs for each regression reported in the main text as well as equivalent models with a log-transformed dependent variable (which often yielded more normally distributed residual estimates, but showed no qualitative differences from the results reported in the main text).

Table 2

Linear regression output for the analysis of proportion photos showing child object handling by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.08	0.03	2.69	0.01
Age	0.00	0.00	2.80	0.01
SiteMayan	-0.06	0.04	-1.49	0.14
Age:SiteMayan	0.00	0.00	2.32	0.02

Table 3

Linear regression output for the analysis of log-transformed proportion photos showing child object handling by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	-3.01	0.24	-12.52	0.00
Age	0.04	0.01	3.81	0.00
SiteMayan	-0.53	0.35	-1.53	0.13
Age:SiteMayan	0.03	0.01	2.02	0.05

Table 4

Linear regression output for the analysis of child object handling burstiness by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.48	0.05	9.53	0.00
Age	0.00	0.00	0.00	1.00
SiteMayan	-0.08	0.08	-0.97	0.34
Age:SiteMayan	0.00	0.00	0.03	0.97

Table 5

Linear regression output for the analysis of log-transformed child object handling burstiness by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	-0.92	0.17	-5.28	0.00
Age	0.01	0.01	1.27	0.21
SiteMayan	0.13	0.27	0.51	0.61
Age:SiteMayan	-0.01	0.01	-1.09	0.28

Table 6

Linear regression output for the analysis of proportion photos showing child-relevant interactant object handling by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.03	0.01	4.83	0.00
Age	0.00	0.00	-1.76	0.08
SiteMayan	0.00	0.01	0.54	0.59
Age:SiteMayan	0.00	0.00	-0.96	0.34

Table 7

Linear regression output for the analysis of log-transformed proportion photos showing child-relevant interactant object handling by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	-3.52	0.16	-21.40	0.00
Age	-0.01	0.01	-1.67	0.10
SiteMayan	0.26	0.24	1.12	0.27
Age:SiteMayan	-0.02	0.01	-1.63	0.11

Table 8

Linear regression output for the analysis of child-relevant interactant object handling burstiness by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.35	0.06	5.77	0.00
Age	0.00	0.00	-1.91	0.06
SiteMayan	0.04	0.10	0.40	0.69
Age:SiteMayan	-0.01	0.00	-2.11	0.04

Table 9

Linear regression output for the analysis of log-transformed child-relevant interactant object handling burstiness by child age, population, and their interaction.

	B	SE	t	p
(Intercept)	0.25	0.16	1.58	0.12
Age	0.00	0.01	-0.59	0.56
SiteMayan	0.33	0.25	1.31	0.20
Age:SiteMayan	-0.03	0.01	-2.65	0.01