Cognition 232 (2023) 105326

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Short communication

Show or tell? Exploring when (and why) teaching with language outperforms demonstration

Theodore R. Sumers [a,*], Mark K. Ho [a], Robert D. Hawkins [b], Thomas L. Griffiths [a,b]

[a] Department of Computer Science, Princeton University, Princeton, NJ, United States of America
[b] Department of Psychology, Princeton University, Princeton, NJ, United States of America

ARTICLE INFO

Keywords:
Pedagogy
Communication
Language
Demonstration
Abstraction

ABSTRACT

People use a wide range of communicative acts across different modalities, from concrete demonstrations to abstract language. While these modalities are typically studied independently, we take a comparative approach and ask *when* and *why* one modality might outperform another. We present a series of real-time, multi-player experiments asking participants to teach concepts using either demonstrations or language. Our first experiment (N = 416) asks *when* language might outperform demonstration. We manipulate the *complexity* of the concept being taught and find that language communicates complex concepts more effectively than demonstration. We then ask *why* language succeeds in this setting. We hypothesized that language allowed teachers to reference *abstract* object features (e.g., shapes and colors), while demonstration teachers could only provide concrete examples (specific positive or negative objects). To test this hypothesis, our second experiment (N = 568) ablated object features from the teacher's interface. This manipulation severely impaired linguistic (but not demonstrative) teaching. Our findings suggest that language communicates complex concepts by directly transmitting abstract rules. In contrast, demonstrations transmit examples, requiring the learner to infer the rules.

1. Introduction

Human teaching takes many forms. For instance, imagine teaching a friend how to play chess and trying to convey *how rooks move*. You could *demonstrate* the rule by dragging a rook along a row, then a column. Alternatively, you could *describe* the rule with language, saying "Rooks move along rows and columns" without ever touching a piece. These distinct modalities both play a role in human pedagogy—but what are their strengths, and how do they work? When and why is language more effective than demonstration?

Researchers have taken different approaches to this question. Classic work has provided rich qualitative insights into natural settings (Carroll & Bandura, 1990; Chi, 2013; Scribner & Cole, 1973). For example, younger children tend to teach via nonverbal demonstration, while older children and adults tend to use language (Ellis & Rogoff, 1982; Strauss, Ziv, & Stein, 2002). Meanwhile, more tightly controlled experiments have tested quantitative predictions from computational theories (e.g., Shafto, Goodman, & Griffiths, 2014), providing mechanistic insight into specific behaviors such as demonstration (Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Gweon, Tenenbaum, & Schulz, 2010; Ho, Cushman, Littman, & Austerweil, 2021), verbal explanation (Chopra, Tessler, & Goodman, 2019), or evaluative feedback (Ho, Cushman, Littman, & Austerweil, 2019). Yet this work has largely focused on *individual* teaching modalities without considering relationships between them (but see Tessler, Bridgers, & Tenenbaum, 2020). This paper seeks a middle ground, introducing a novel interactive teaching paradigm to directly compare modalities while controlling for factors that may mediate their efficacy.

We report a pair of studies that compare two communication modalities—*demonstration* and *language*. Demonstration (*showing* someone how chess pieces move) is a form of example-based teaching (Bandura, 1965; Csibra & Gergely, 2009; Shneidman, Gweon, Schulz, & Woodward, 2016). Demonstrations are non-symbolic communicative acts, relying on shared understanding of a problem domain (such as physical affordances and goals; Gergely & Csibra, 2003; Gergely & Jacob, 2012). The capacity to generate and understand demonstrations is foundational, emerging early in childhood (Butler & Markman, 2012; Gweon & Schulz, 2018; Gweon et al., 2010; Király, Csibra, & Gergely, 2013; Leonard, Lee, & Schulz, 2017). Language (e.g., *telling* someone the rules of chess) is another cognitively privileged mechanism for transmitting knowledge (Grice, 1975). Language relies on shared conventions (Clark, 1996) to convey information about categories (Gelman & Markman, 1986, 1987), relations (Loewenstein & Gentner,

* Corresponding author.
E-mail address: sumers@princeton.edu (T.R. Sumers).

https://doi.org/10.1016/j.cognition.2022.105326
Received 5 November 2021; Received in revised form 6 June 2022; Accepted 10 November 2022
Available online 5 December 2022

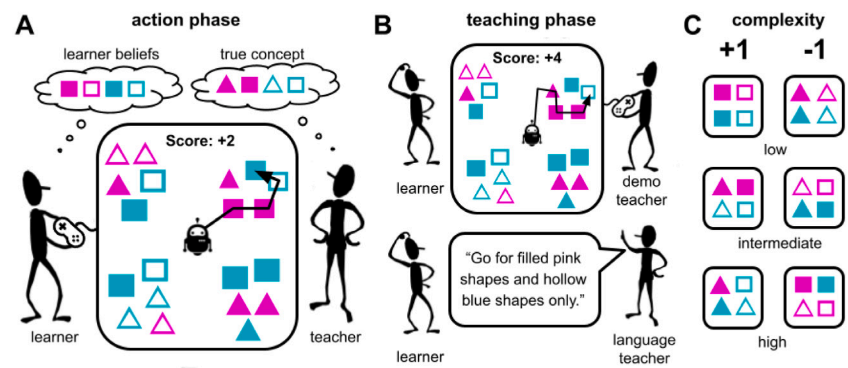

Fig. 1. *Experimental paradigm and conditions*. A: Each trial began with an action phase, in which the learner was placed in a new level and collected objects according to their beliefs about which were positive. Learners generally began the task choosing randomly (trial 1 score: M = −.09, SD = 2.23). B: After the action phase, teachers communicated via demonstration or language. Over the course of the experiment (10 trials), teachers were generally able to communicate the concept (trial 10 score: M = 3.30, SD = 2.65), although performance varied substantially by condition (Table S1). C: Manipulating concept complexity. Positive and negative objects were "masked" by Boolean concepts of varying complexity.

2005; Rattermann & Gentner, 1998), and causal structures (Lombrozo, 2006; Tessler & Goodman, 2019). While both are fundamental means of knowledge transmission, they operate by different mechanisms: language uses conventionalized forms of reference to transmit abstract information, while demonstrations instantiate that information in concrete examples (Csibra & Shamsudheen, 2015).

But *when* and *why* might one be preferred? We develop a pair of concept-teaching studies to address these questions. Experiment 1 asks *when* language outperforms demonstration, varying the teaching *content* by manipulating the concept complexity (Feldman, 2000; Shepard, Hovland, & Jenkins, 1961). We find that concept complexity has systematic and differential effects on these modalities: language outperforms demonstration when communicating complex concepts. We then ask *why*, hypothesizing that language succeeds by encoding the abstract, underlying rule (Lombrozo, 2006). Experiment 2 tests this hypothesis by varying the teaching *context*. We ablate the teacher's view of object features, while leaving spatial context intact, effectively forcing teachers to communicate about specific objects rather than abstract object features. We find that this manipulation severely impairs linguistic teaching while leaving demonstrative teaching unaffected. This systematic comparison supports the idea that language relies on shared abstractions to efficiently transmit complex concepts. In contrast, demonstrations struggle to communicate such concepts, but are less reliant on shared abstractions.

2. Experiment 1: Linguistic teaching conveys complex concepts better than demonstrations when shared abstractions are available

Concepts can be more or less complex. Intuitively, the rule for *how rooks move* is simpler than the rule for *castling*.[1] Increasing complexity makes individual learning more difficult (Feldman, 2000; Shepard et al., 1961) and favors abstract problem representations (Koedinger, Alibali, & Nathan, 2008). We thus predicted that both modalities would suffice for teaching simple concepts, but language could better convey complex ones. To test this prediction, we introduce a collaborative teaching game that places pairs of participants in an interactive virtual environment. The environment contains colored shapes with different point values. One participant, the *teacher*, can see the objects' values but is unable to act in the environment. Consequently, they must communicate this information to their partner, the *learner*, who controls an avatar in the environment. We manipulated the communication modality and concept complexity in a fully between-subjects design. In the *language* condition, teachers sent chat messages to the learner; in the *demonstration* condition, teachers produced movement trajectories instead (see Fig. 1). Finally, we manipulated complexity by changing the mapping of point values to object visual features, giving rise to different underlying concepts.[2]

2.1. Methods

2.1.1. Participants

We recruited 480 participants using Prolific (www.prolific.co). Participants were required to be fluent in English, possess an approval rating of 95%, and be located within the United States, United Kingdom, or Ireland. They were paid $2.25 with a score-based bonus up to $1.75. We excluded seven participants who accidentally re-started the experiment after refreshing their browser, for a final sample of 227 pairs, 208 of whom completed all 10 levels.

2.1.2. Procedure

After reading instructions, all participants completed two practice trials to familiarize them with the environment. Participants were then paired, assigned teacher or learner roles, and played a practice trial together. They then played 10 trials of a communication task. Each trial proceeded as follows. First, in the action phase (Fig. 1A), the learner was given 30 s to study the level and plan their movements. They then had 8 s to move the avatar to collect objects while the teacher watched. The learner was only shown the shapes and colors, but the teacher could also see the underlying values. After 8 s, the action phase ended and both players were shown the net score.[3] They then proceeded to a teaching phase (Fig. 1B). *Language* teachers were allowed to message their partner through a one-way chat interface; *demonstration* teachers re-played the level while the learner watched. After 10 trials, both players completed an exit survey, which included a measure of concept knowledge: learners were asked the value of each of the 8 possible objects (+1/-1/"Don't Know").

[1] "A king can be moved two spaces towards a rook, and the rook to the opposite side of the king, *if* the king is not in check, neither piece has previously moved, there are no pieces between them, and the king does not pass through a square that would place it in check."

[2] The study was approved by the Princeton IRB; our sample size and planned analyses were pre-registered at https://aspredicted.org/S8S_FMQ. Code and data are available at https://github.com/tsumers/show-or-tell.

[3] For example, collecting three objects with a value of +1 and one with a value of −1 would yield a net score of 2.

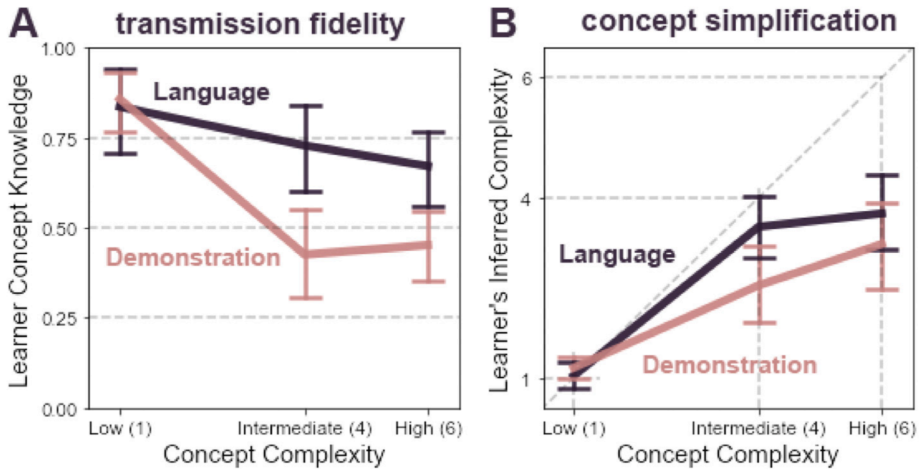

Fig. 2. *Experiment 1 results.* A: linguistic teaching sustained higher transmission fidelity as concept complexity increased. B: learners tended to simplify the concept being taught, inferring a lower-complexity concept. This effect was particularly pronounced for demonstration learners. Error bars show 95% CI.

2.1.3. Stimuli

The player began each trial in the center of the screen, with a cluster of five objects in each corner. Each object was assigned a positive (+1) or negative (−1) value and a set of visual features according to different Boolean concepts (Fig. 1C). We used three features, each taking two possible values, for a total of $2^3 = 8$ feature combinations (color: blue or pink; shape: square or triangle; fill: solid or hollow). We assigned features using three levels of complexity — *low*, *intermediate*, and *high*, corresponding to Shepard concepts I, II, and III (Shepard et al., 1961) or Boolean complexity 1, 4, and 6, in the taxonomy of Feldman (2000). Within each complexity condition, we randomized features to mitigate salience biases.[4] To allow direct comparison of movement trajectories, values of objects in each location were generated randomly but fixed across participants for each trial. Thus all participants proceeded through the same set of 10 randomly generated maps, but with different visual features assigned to locations.

2.2. Results

2.2.1. Transmission fidelity

Our primary hypothesis concerned transmission fidelity. We expected that linguistic teachers could convey the underlying concept more effectively than demonstrative teachers, especially for more complex concepts. We operationalized transmission fidelity using *learner concept accuracy*, the number of correct responses in the exit survey (Fig. 2A). To test our hypothesis, we used a binomial logistic regression predicting the number of correct responses (out of 8). We included predictors for communication modality, (Boolean) concept complexity, and their interaction. To interpret main effects, we centered the regressors.[5] Language afforded significantly higher transmission fidelity than demonstration ($\beta = .38$, $t(202) = 6.66$, $p < .0001$). As complexity increased, fidelity in both decreased ($\beta = -.29$, $t(202) = -9.98$, $p < .0001$). However, we also found a significant interaction. On more complex concepts, demonstration fared significantly worse than language ($\beta = .11$, $t(202) = 3.74$, $p < .001$; see Table S2). Task performance followed a similar pattern (see Appendix A).

2.2.2. Concept simplification

Our analysis of transmission fidelity asked *how much* concept knowledge learners acquired. We now investigate *what* they learned. Rather than make random errors, we hypothesized that demonstration learners would formulate a simpler rule based on fewer features (Goodman, Tenenbaum, Feldman, & Griffiths, 2008), thereby reducing its complexity. For example, a demonstration learner assigned the most complex concept in Fig. 1C (Boolean complexity 6) might observe their teacher collect mostly triangles and infer that all and only triangles were positive (Boolean complexity 1). To operationalize this prediction, we compared the Boolean complexity of the learner's survey responses to the true concept (Fig. 2B). We ran a regression analysis predicting the complexity of the learner's inferred concept as a function of the communication modality (contrast-coded as before), controlling for true complexity (dummy-coded with *low* as the baseline). We found a weak effect of modality ($\beta = .24$, $t(202) = 2.01$, $p = .045$; see Table S4 for full model). While all teachers struggled to convey the most complex concepts, demonstration learners simplified even intermediate ones.

2.3. Discussion

As predicted, linguistic teaching outperformed demonstration at transmitting complex concepts. Demonstration learners tended to prefer simple explanatory rules; in contrast, language allowed teachers to more precisely specify complex feature combinations. Of course, factors aside from modality may inhibit learning of more complex concepts. For example, teachers may have deliberately simplified harder concepts, teaching "good enough" rules intended to maximize the learner score rather than transmit the full concept (see Table S6 for an example of teaching a simplified concept first). Learning may also be affected by more general factors such as memory and attention. However, because these factors were held constant across conditions, our results reflect modality-based differences for the communicative pair as a whole.

These findings suggest that when teachers know the abstract features that define a concept, language matches or outperforms demonstration. But *how* does language transmit such complex concepts? We hypothesized that language succeeds by allowing direct reference to these task-relevant abstractions (Loewenstein & Gentner, 2005; Rattermann & Gentner, 1998). Our second experiment ablates these features, where we predict that language—but not demonstrations—would be severely impaired.

[4] For example, in the *low* complexity condition, a single dimension determined the value. Pairs could be assigned blue as positive (pink as negative), pink as positive (blue as negative), squares as positive (triangles as negative), and so on.

[5] We treated complexity as a continuous variable and subtracted the mean, $(1 + 4 + 6)/3$. We sum coded the communication condition (*language* = 1, *demonstration* = −1).

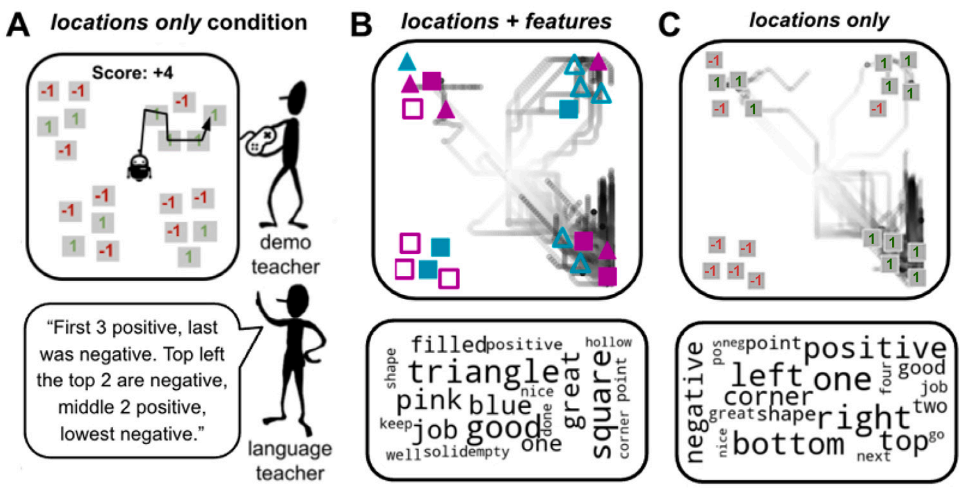

Fig. 3. *Manipulating access to abstractions in Experiment 2.* A: Under the *locations only* condition, teachers could not see object features. They could only see *which* objects were positive. B/C: *Locations only* affected linguistic teaching, but not demonstrations. Top: demonstration trajectories from the second level. Teachers in both conditions relied on score-maximizing demonstrations; here, they overwhelmingly collected the five positive objects in the bottom right, then moved as far as possible towards the top right, which was the next best corner. Bottom: most common words in each condition. Linguistic teachers in the two conditions used very different strategies: *Locations and features* teachers relied primarily on the features (shape, color, and fill), while *Locations only* teachers relied on behavioral ("first 3 positive") or spatial ("top left the top 2 are negative") references to describe which objects were positive.

3. Experiment 2: Linguistic (but not demonstrative) teaching relies on shared abstractions

Experiment 1 found that language outperforms demonstration at communicating complex concepts. Why might this be the case? Consider the statement "Rooks move along rows and columns." *Rook* is an *abstraction*: it refers to a *set* of objects defined by the label *rook* (Gelman, 2004; Gelman & Markman, 1986). But what if one did not have such a label? One could imagine referring to *specific* rooks: "*That piece in the corner* moves along rows and columns." This reference creates ambiguity when it comes time for the listener to generalize. Which other pieces does this same property extend to? What about the opponent's identically-shaped (but differently colored) pieces? Intuitively, such reasoning seems cumbersome. Indeed, in Experiment 1, most language teachers expressed rules by referring to abstract object *features*: "filled pink shapes and hollow blue shapes" (Fig. 1B), rather than specific objects. In contrast, demonstrations *cannot* refer to abstract features: physical actions necessarily engage with physical objects.

This observation motivates our hypothesis: such *shared abstractions* underpin language's success in Experiment 1. To test this hypothesis, we introduced a new manipulation: we removed the abstract object features (shape, color, and fill) from the teacher's interface. Teachers in the "Locations and Features" condition had the same interface as Experiment 1, while teachers in the "Locations Only" condition could only see objects' values (Fig. 3A). Such teachers knew which objects were positive or negative but could not refer to their abstract features. We hypothesized this ablation would impair linguistic teaching while leaving demonstrative teaching largely unaffected.[6]

3.1. Methods

3.1.1. Participants

We recruited a (pre-registered) sample of 640 participants on Prolific using the same qualifications and payment as Experiment 1. Thirteen were excluded after accidentally refreshing their browser and re-starting the task. Our final sample contained 298 pairs, of whom 284 completed all 10 levels.

3.1.2. Materials and procedure

We again used a between-subjects design crossing modality (*language* vs. *demonstration*) with complexity (*low* vs. *intermediate*; the hardest concept level was removed). To manipulate teachers' ability to use abstractions, we introduced a third condition, shared context (*features and locations* vs. *locations only*). *Features and locations* teachers were given the same user interface as Experiment 1. *Locations only* teachers were given an ablated interface: they saw the value of objects, but *not* their features (shape, color, and fill; Fig. 3A). Otherwise, the procedure was the same as Experiment 1, using the same sequence of 10 trials.

3.2. Results

3.2.1. Transmission fidelity

As in Experiment 1, our primary variable of interest was the learner's concept knowledge (Fig. 4A). We hypothesized *locations only* context would negatively affect linguistic teaching, without impairing demonstration. To test for an interaction between communication modality and context, we again used a binomial logistic regression predicting correct responses on the learner's exit survey. Predictors included complexity, communication modality, context (all effect-coded), and the communication–context interaction. All four terms were significant, critically including the communication–context interaction ($\beta = .34, t(279) = 6.90, p < .001$; see Table S9), consistent with our hypothesis that *locations only* context would impair linguistic teaching more than demonstration (see Appendix B for full regression models).

3.2.2. Cumulative scores

Task performance may provide another perspective on communicative success: did different teaching modalities allow learners to reach higher scores? Specifically, we examine the *cumulative score* (Fig. 4B). If two learners end the experiment with equal concept knowledge but different scores, this implies that the higher-scoring learner acquired the concept *faster*. We used a linear regression to predict each learner's cumulative score using same predictors as above.[7] The results suggest even more extreme limitations of *locations only* context linguistic teaching. Even in the *low* complexity condition, these pairs scored worse than

[6] Our pre-registration is available at https://aspredicted.org/ZHS_YW1.

[7] This model was a simplification of our pre-registered analysis; see Appendix E for details.

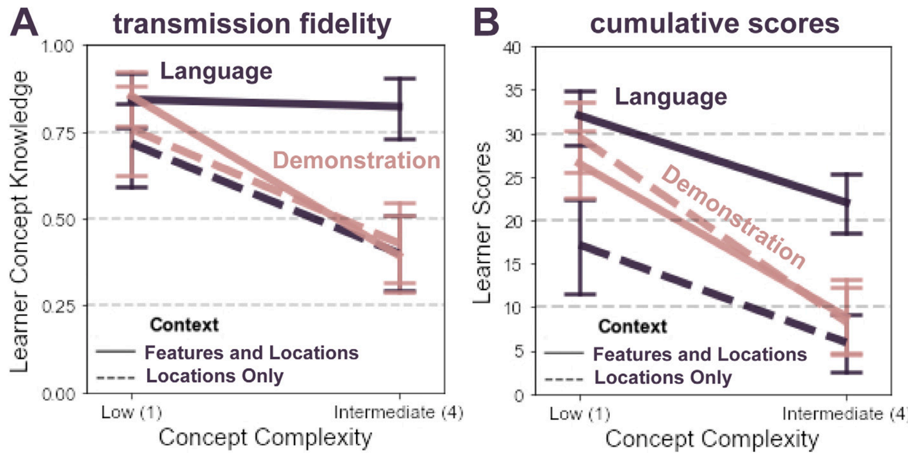

Fig. 4. *Experiment 2 results.* A: All modalities succeeded at communicating low complexity concepts. However, increased complexity severely impairs transmission fidelity for all *demonstrations* and *locations only* linguistic teachers—while leaving linguistic teachers with access to abstract features largely unaffected. B: Cumulative scores reveal that *location-only* linguistic teaching lagged behind other conditions (see Figure S4 for per-level scores). Error bars show 95% CI.

pairs in either *demonstration* condition (interaction $\beta = 4.13, t(279) = 5.58$, $p < .001$, see Fig. S4 and Table S10 for full model).

3.2.3. Teacher strategies and subjective difficulty

How did removing object features affect teaching? Language teachers in different context conditions used very different words (Fig. 3B/C; Appendix C). *Features and locations* teachers relied heavily on the abstract features (color, shape, and fill), while *locations only* teachers primarily referenced specific positive objects (e.g., using spatial references).

In contrast, demonstration teachers were largely unaffected by context. Teachers' most common strategy was to maximize their score: 75% (1078 of 1430) of demonstrations achieved the highest possible points per level. There was not a significant difference in teachers' rates of score-maximizing demonstrations across the four conditions ($F(3, 139) = 1.515, p = .213$). In other words, most demonstration teachers taught by acting optimally—a viable strategy even when the teacher and learner have differing representations of the environment.

Finally, we explored the division of labor across conditions and roles by asking participants to rate how difficult the game was (Fig. 5). Participants generally found the *low* complexity condition easy. However, at *intermediate* complexity, theoretically interesting differences emerged across conditions and roles ($F(7, 278) = 18.76, p < .0001$; see Table S21 for Tukey HSD tests). *Language* teachers and learners agreed that sharing *features and locations* was easier than *locations only*. Both roles thought *features and locations* was easy; both thought *locations only* was hard. In contrast, in *demonstration* conditions, *learners* found the task more difficult than *teachers*; the shared *context* did not matter. This asymmetry suggests that demonstrative teaching places the cognitive burden on the learner: it was easy for teachers to provide demonstrations, but difficult for learners to infer the underlying concept.

3.3. Discussion

Our findings support the hypothesis that language succeeded in Experiment 1 by relying on shared *abstractions*: allowing reference to object *features* rather than the objects themselves. Lacking access to these features, linguistic teaching suffered while demonstrations did not. We note that even with *locations only*, linguistic teachers were in principle capable of communicating more information than demonstrative teachers. Indeed, one *locations only* language teacher specified the value of every object in every trial (Table S18). In general, however, such strategies were difficult and inefficient: despite sending more text (Fig. S2), *locations only* teachers struggled to transmit the concept.[8]

4. General discussion

Teaching is central to human social life and has been examined extensively from evolutionary (Tomasello, 2009), developmental (Csibra & Gergely, 2009), cross-cultural (Shneidman et al., 2016) and computational (Shafto et al., 2014) perspectives. Yet recent experimental work has largely studied single modalities without considering their relative efficacy. Our study compared different modalities in a rich but controlled communication task.

In our experiments, *language* teachers excelled at communicating abstract information. However, without access to object features, they found it challenging to communicate the same information in terms of specific objects. In contrast, *demonstration* teachers struggled to communicate complex concepts, but were largely unaffected by removing object features. This pattern highlights a key difference between modalities: language may use *abstract* labels to convey *abstract* information, whereas demonstrations use *concrete* examples, forcing the learner to *infer* abstract information (Butler & Markman, 2016; Hume, 1748). While humans possess a remarkable ability to learn from such data (Baldwin, Markman, & Melartin, 1993; Butler & Markman, 2012; Csibra & Gergely, 2009; Gergely, Egyed, & Király, 2007; Shafto et al., 2014), a conventionalized lexicon allows language to directly encode abstract information (Csibra & Shamsudheen, 2015; Gelman & Markman, 1986). Notably, while learning from expert demonstrations is a historically popular approach in robotics (Abbeel & Ng, 2004; Argall, Chernova, Veloso, & Browning, 2009), recent trends in the field emphasize learning from language instead (Luketina et al., 2019; Tellex, Gopalan, Kress-Gazit, & Matuszek, 2020).

However, the present study captures only a small slice of human communication. We enforced one-way communication, hindering discourse repairs (Dingemanse et al., 2015) and *ad-hoc* formation of new linguistic abstractions (McCarthy, Hawkins, Wang, Holdaway, & Fan, 2021) or signaling conventions (Galantucci & Garrod,

[8] We note two possible concerns regarding *locations only* linguistic teaching. First, the condition may have been difficult for teachers to understand. Examination of chat logs revealed several teachers who either failed to understand how the learner's perspective differed from theirs or offered encouragement rather than teaching. Removing these pairs, however, did not affect our results (Appendix D). Second, the lack of spatial structure made referencing specific objects particularly challenging. However, a followup study with more spatial structure replicated our findings (Appendix F).

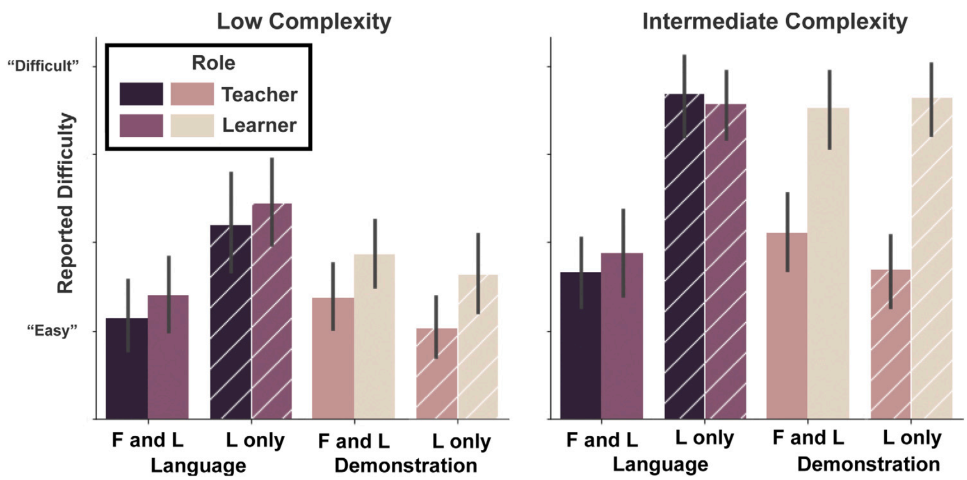

Fig. 5. *Participants' self-reported difficulty ratings.* F = features, L = Locations. Participants in the *low* complexity condition generally found the experiment easy. However, significant differences emerged in the *intermediate* complexity condition. In the language conditions, difficulty varied by *context*: teachers and learners agreed that the *features and locations* condition was easy and the *locations only* condition was hard. In the demonstration conditions, difficulty varied by *role*: teachers thought the task was easy, while learners thought the task was hard. Error bars show 95% CI.

2011; Scott-Phillips, Kirby, & Ritchie, 2009). In addition, we studied transmission of a particular kind of information: Boolean concepts (Shepard et al., 1961). Tradeoffs between modalities may differ when teaching other types of concepts (e.g., fundamentally physical athletic concepts, Martens, 1975). Indeed, despite extensive specialized vocabularies, demonstrative teaching is widespread in domains such as music (Dickey, 1991; Weeks, 1996), dance (Keevallik, 2010), cooking (Mondada, 2014), sports (Evans & Reynolds, 2016), martial arts (Råman, 2019), medical procedures (Svensson, Luff, & Heath, 2009) and vocational training (Asplund & Kilbrink, 2018). Finally, we studied modalities independently; naturalistic teaching may interleave them, such as narrating a demonstration (Carroll & Bandura, 1990) to establish shared linguistic abstractions.

Our findings suggest several avenues of future work. First, our experimental paradigm may support extending computational models to interactive multi-modal teaching. Communication is classically viewed as a means to intervene on abstract beliefs (Goodman & Frank, 2016; Grice, 1975; Shafto, Wang, & Wang, 2021), but it is inherently grounded in real-world actions: humans necessarily infer partners' beliefs through observations (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017) and integrate information from different modalities (Butler & Tomasello, 2016; Carroll & Bandura, 1990; Davis-Unger & Carlson, 2008). Thus, models of communication situated within decision-making environments (Bridgers, Jara-Ettinger, & Gweon, 2019; Fisac et al., 2020; Sumers, Hawkins, Ho, & Griffiths, 2021) could begin to explain interactive, multi-modal communicative acts.

Finally, transmission fidelity is central to accounts of cumulative cultural evolution (Lewis & Laland, 2012; Tomasello, 2009), but the question of which modalities are necessary or sufficient to accumulate knowledge remains open. Empirical work has used tasks such as stone tool making (Lombao, Guardiola, & Mosquera, 2017; Morgan, Uomini, Rendell, Chouinard-Thuly, Street, Lewis, Cross, Evans, Kearney, de la Torre, Whiten, & Laland, 2015) or complex problem solving (Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Tessler, Tsividis, Madeano, Harper, & Tenenbaum, 2021; Thompson, van Opheusden, Sumers, & Griffiths, 2022), but measuring knowledge accumulation in these domains is challenging. In contrast, Boolean concepts allow us to precisely track *what* individuals learned.[9] Future work could extend our paradigm to iterated settings to directly measure the acquisition and refinement of complex concepts over multiple generations.

CRediT authorship contribution statement

Theodore R. Sumers: Conceptualization, Software, Formal analysis, Investigation, Data curation, Visualization, Funding acquisition, Writing. **Mark K. Ho:** Conceptualization, Formal analysis, Supervision, Writing. **Robert D. Hawkins:** Conceptualization, Formal analysis, Supervision, Writing. **Thomas L. Griffiths:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors declare that there is no conflict of interest. TRS is supported by the NDSEG Fellowship Program and RDH is supported by the National Science Foundation (grant #1911835). This work was additionally supported by the National Science Foundation (grant number #1544924) and the John Templeton Foundation (grant #61454). We thank M.H. Tessler and our anonymous reviewers for insightful and constructive feedback.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105326.

[9] For example, Lombao et al. (2017) use observed behaviors (coded sequences of actions) and task outcomes (the quality and quantity of stone flakes produced) to compare teaching modalities. They then suggest that language succeeds due to better communication of specific conceptual knowledge (e.g. the idea of a "percussion platform"). Our paradigm allows us to query the learner for their precise concept knowledge, providing direct evidence for effects such as concept simplification (Experiment 1, Fig. 2B).

References

Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the 21st International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1015330.1015430.

Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, *57*(5), 469–483. http://dx.doi.org/10.1016/j.robot.2008.10.024.

Asplund, S.-B., & Kilbrink, N. (2018). Learning how (and how not) to weld: Vocational learning in technical vocational education. *Scandinavian Journal of Educational Research*, *62*(1), 1–16. http://dx.doi.org/10.1080/00313831.2016.1188147.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behavior*, *1*, 0064. http://dx.doi.org/10.1038/s41562-017-0064.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. *Child Development*, *64*(3), 711. http://dx.doi.org/10.2307/1131213.

Bandura, A. (1965). Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *vol. 2*, *Advances in experimental social psychology* (pp. 1–55). Academic Press, http://dx.doi.org/10.1016/S0065-2601(08)60102-1.

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2019). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behavior*, *4*(2), 144–152. http://dx.doi.org/10.1038/s41562-019-0748-6.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*(3), 331–340. http://dx.doi.org/10.1016/j.cognition.2010.12.001.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416–1428. http://dx.doi.org/10.1111/j.1467-8624.2012.01775.x.

Butler, L. P., & Markman, E. M. (2016). Navigating pedagogy: Children's developing capacities for learning from pedagogical interactions. *Cognitive Development*, *38*, 27–35. http://dx.doi.org/10.1016/j.cogdev.2016.01.001.

Butler, L. P., & Tomasello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology*, *145*, 64–78. http://dx.doi.org/10.1016/j.jecp.2015.12.001.

Carroll, W. R., & Bandura, A. (1990). Representational guidance of action production in observational learning: A causal analysis. *Journal of Motor Behavior*, *22*(1), 85–97. http://dx.doi.org/10.1080/00222895.1990.10735503.

Chi, M. T. (2013). Learning from observing an expert's demonstration, explanations, and dialogues. *Expertise and Skill Acquisition: The Impact of William G. Chase*, *1*, 1–27. http://dx.doi.org/10.4324/9780203074541.

Chopra, S., Tessler, M. H., & Goodman, N. (2019). The first crank of the cultural ratchet: Learning and transmitting concepts through language. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 226–232).

Clark, H. H. (1996). *Using language*. Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511620539.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. http://dx.doi.org/10.1016/j.tics.2009.01.005.

Csibra, G., & Shamsudheen, R. (2015). Nonverbal generics: Human infants interpret objects as symbols of object kinds. *Annual Review of Psychology*, *66*(1), 689–710. http://dx.doi.org/10.1146/annurev-psych-010814-015232.

Davis-Unger, A. C., & Carlson, S. M. (2008). Children's teaching skills: The role of theory of mind and executive function. *Mind, Brain, and Education*, *2*(3), 128–135. http://dx.doi.org/10.1111/j.1751-228X.2008.00043.x.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *335*(6072), 1114–1118. http://dx.doi.org/10.1126/science.1213969.

Dickey, M. R. (1991). A comparison of verbal instruction and nonverbal teacher-student modeling in instrumental ensembles. *Journal of Research in Music Education*, *39*(2), 132–142. http://dx.doi.org/10.2307/3344693.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal principles in the repair of communication problems. *PLoS One*, *10*(9), Article e0136100. http://dx.doi.org/10.1371/journal.pone.0136100.

Ellis, S., & Rogoff, B. (1982). The strategies and efficacy of child versus adult teachers. *Child Development*, *53*, 730–735. http://dx.doi.org/10.2307/1129386.

Evans, B., & Reynolds, E. (2016). The organization of corrective demonstrations using embodied action in sports coaching feedback. *Symbolic Interaction*, *39*(4), 525–556. http://dx.doi.org/10.1002/SYMB.255.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633. http://dx.doi.org/10.1038/35036586.

Fisac, J. F., Gates, M. A., Hamrick, J. B., Liu, C., Hadfield-Menell, D., Palaniappan, M., et al. (2020). Pragmatic-pedagogic value alignment. In *Robotics Research* (pp. 49–57). Springer.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: A review. *Frontiers in Human Neuroscience*, *5*, http://dx.doi.org/10.3389/fnhum.2011.00011.

Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409. http://dx.doi.org/10.1016/j.tics.2004.07.001.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. http://dx.doi.org/10.1016/0010-0277(86)90034-x.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, *58*(6), 1532. http://dx.doi.org/10.2307/1130693.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. http://dx.doi.org/10.1016/s1364-6613(03)00128-1.

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, *10*(1), 139–146. http://dx.doi.org/10.1111/j.1467-7687.2007.00576.x.

Gergely, G., & Jacob, P. (2012). Reasoning about instrumental and communicative agency in human infancy. In *Rational constructivism in cognitive development* (pp. 59–94). Elsevier, http://dx.doi.org/10.1016/b978-0-12-397919-3.00003-4.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. http://dx.doi.org/10.1016/j.tics.2016.08.005.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154. http://dx.doi.org/10.1080/03640210701802071.

Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics: Vol. 3: Speech acts* (pp. 41–58). New York: Academic Press.

Gweon, H., & Schulz, L. (2018). From exploration to instruction: Children learn from exploration and tailor their demonstrations to observers' goals and competence. *Child Development*, *90*(1), http://dx.doi.org/10.1111/cdev.13059.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071. http://dx.doi.org/10.1073/pnas.1003095107.

Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2019). People teach with rewards and punishments as communication, not reinforcements. *Journal of Experimental Psychology: General*, *148*(3), 520–549. http://dx.doi.org/10.1037/xge0000569.

Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2021). Communication in action: Planning and interpreting communicative demonstrations. *Journal of Experimental Psychology: General*, *150*(11), 2246–2272. http://dx.doi.org/10.1037/xge0001035.

Hume, D. (1748). *An Enquiry concerning human understanding*.

Keevallik, L. (2010). Bodily quoting in dance correction. *Research on Language and Social Interaction*, *43*(4), 401–426. http://dx.doi.org/10.1080/08351813.2010.518065.

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology*, *116*, 471–486. http://dx.doi.org/10.1016/j.jecp.2012.12.003.

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science*, *32*(2), 366–397. http://dx.doi.org/10.1080/03640210701863933.

Leonard, J. A., Lee, Y., & Schulz, L. E. (2017). Infants make more attempts to achieve a goal when they see adults persist. *Science*, *357*(6357), 1290–1294. http://dx.doi.org/10.1126/science.aan2317.

Lewis, H., & Laland, K. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society of London. Series B*, *367*(1599), 2171–2180. http://dx.doi.org/10.1098/rstb.2012.0119.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. http://dx.doi.org/10.1016/j.cogpsych.2004.09.004.

Lombao, D., Guardiola, M., & Mosquera, M. (2017). Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. *Scientific Reports*, *7*(1), http://dx.doi.org/10.1038/s41598-017-14322-y.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. http://dx.doi.org/10.1016/j.tics.2006.08.004.

Luketina, J., Nardelli, N., Farquhar, G., Foerster, J., Andreas, J., Grefenstette, E., et al. (2019). A survey of reinforcement learning informed by natural language. In *International joint conferences on artificial intelligence organization*. http://dx.doi.org/10.24963/ijcai.2019/880.

Martens, R. (1975). *Social psychology and physical activity*. Longman Higher Education.

McCarthy, W. P., Hawkins, R. D., Wang, H., Holdaway, C., & Fan, J. E. (2021). Learning to communicate about shared procedural abstractions. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*.

Mondada, L. (2014). Cooking instructions and the shaping of things in the kitchen. *Interacting with Objects: Language, Materiality, and Social Activity*, 199–226. http://dx.doi.org/10.1075/z.186.09mon.

Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., et al. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, *6*, http://dx.doi.org/10.1038/ncomms7029.

Råman, J. (2019). Budo demonstrations as shared accomplishments: the modalities of guiding in the joint teaching of physical skills. *Journal of Pragmatics*, *150*, 17–38. http://dx.doi.org/10.1016/j.pragma.2019.06.014.

Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In *Advances in analogy research : Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282).

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233.

Scribner, S., & Cole, M. (1973). Cognitive consequences of formal and informal education. *Science*, *182*(4112), 553–559. http://dx.doi.org/10.1126/science.182.4112.553.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89. http://dx.doi.org/10.1016/j.cogpsych.2013.12.004.

Shafto, P., Wang, J., & Wang, P. (2021). Cooperative communication as belief transport. *Trends in Cognitive Sciences*, *25*(10), 826–828. http://dx.doi.org/10.1016/j.tics.2021.07.012.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications.. *Psychological Monographs: General and Applied*, *75*(13), 1–42. http://dx.doi.org/10.1037/h0093825.

Shneidman, L., Gweon, H., Schulz, L. E., & Woodward, A. L. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development*, *87*(3), 723–735. http://dx.doi.org/10.1111/cdev.12502.

Strauss, S., Ziv, M., & Stein, A. (2002). Teaching as a natural cognition and its relations to preschoolers' developing theory of mind. *Cognitive Development*, *17*(3–4), 1473–1487. http://dx.doi.org/10.1016/S0885-2014(02)00128-4.

Sumers, T. R., Hawkins, R. D., Ho, M. K., & Griffiths, T. L. (2021). Extending rational models of communication from beliefs to actions. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*.

Svensson, M. S., Luff, P., & Heath, C. (2009). Embedding instruction in practice: contingency and collaboration during surgical training. *Sociology of Health & Illness*, *31*(6), 889–906. http://dx.doi.org/10.1111/j.1467-9566.2009.01195.x.

Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C. (2020). Robots that use language. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*(1), 25–55. http://dx.doi.org/10.1146/annurev-control-101119-071628.

Tessler, M. H., Bridgers, S., & Tenenbaum, J. B. (2020). How many observations is one generic worth? In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, *126*(3), 395–436. http://dx.doi.org/10.1037/rev0000142.

Tessler, M. H., Tsividis, P. A., Madeano, J., Harper, B., & Tenenbaum, J. B. (2021). Growing knowledge culturally across generations to solve novel, complex tasks. In *Proceedings of the 43nd Annual Conference of the Cognitive Science Society*.

Thompson, B., van Opheusden, B., Sumers, T., & Griffiths, T. L. (2022). Complex cognitive algorithms preserved by selective social learning in experimental populations. *Science*, *376*(6588), 95–98. http://dx.doi.org/10.1126/science.abn0915.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard University Press.

Weeks, P. (1996). A rehearsal of a Beethoven passage: An analysis of correction talk. *Research on Language and Social Interaction*, *29*(3), 247–290. http://dx.doi.org/10.1207/s15327973rlsi2903_3.