Title

A formal model of intuitive theories of vision in congenitally blind and sighted adults

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Wang, Ziwen
Musz, Elizabeth
Keil, Sophia
et al.

Publication Date

2024

Intuitive Theory of Vision in Congenitally Blind and Sighted Adults

Ziwen Wang (ZWANG252@Jhu.Edu)
Department of Psychological and Brain Sciences, 3400 N. Charles Street
Baltimore, MD 21218 USA

Elizabeth Musz (EMUSZ1@Jhu.Edu)
Department of Psychological and Brain Sciences, 3400 N. Charles Street
Baltimore, MD 21218 USA

Sophia Keil (SKEIL4@Jhu.Edu)
Department of Psychological and Brain Sciences, 3400 N. Charles Street
Baltimore, MD 21218 USA

Colin Wilson (COLIN.WILSON@Jhu.Edu)
Department of Psychological and Brain Sciences, 3400 N. Charles Street
Baltimore, MD 21218 USA

Marina Bedny (MARINA.BEDNY@Jhu.Edu)
Department of Psychological and Brain Sciences, 3400 N. Charles Street
Baltimore, MD 21218 USA

Abstract

Comparison of visibility inferences across congenitally blind and sighted people provides insight into the contribution of first-person sensory experience to intuitive theories. We hypothesized that both groups understand others' visual experiences via an intuitive theory incorporating variables known to influence visual psychophysics (distance, looking duration, and feature size). Adults born blind (n=20) and sighted (n=40) listened to short scenarios that described an observer looking at another person from different distances and for varying durations. Participants rated how likely the observer would perceive appearance features of the person that varied in size (e.g., eye color vs. hat). A probabilistic formalization of intuitive visibility fit the ratings with high accuracy across scenarios and features. Model parameters were qualitatively identical across groups but blind adults weighted distance and size less. A quantitative and generative intuitive theory of vision develops without first-person sensory access, possibly through linguistic communication, and is fine-tuned by visual experience.

Keywords: blindness; perception; theory of mind; mental model

Introduction

Humans engage in causal and predictive reasoning across diverse cognitive domains, ranging from mechanical causes (intuitive physics) to the invisible mental causes of actions (intuitive psychology) (Gopnik & Meltzoff, 1998; Jara-Ettinger et al., 2016; Ullman et al., 2017; Wellman & Gelman, 1992). This reasoning is governed by mental models known as 'intuitive theories' that encode causal relationships between interconnected concepts (e.g., between beliefs and actions) and are used to make generative inferences in new situations (Carey, 2009; Gopnik, 2003; Wellman & Gelman, 1992; Xu, 2019). A key question that remains unanswered is how such intuitive theories are constructed and modified in response to evidence (Gopnik & Wellman, 2012; Gweon, 2021; Sommerville et al., 2005).

One early-available and important source of evidence is one's own first-person sensory experience. For example, we might learn about gravity by watching objects fall (Kim & Spelke, 1999; Shanon, 1976). Likewise, we could learn that actions are caused by goals and beliefs by introspecting about our own behavior (Sommerville et al., 2005). However, humans are also prodigious social learners. We learn much of what we know not through direct observation, but indirectly, by communicating with others through language (Gweon, 2021; Pinker, 2003; Tooby & DeVore, 1987; Xu, 2019). Recent large language models (LLMs) highlight the incredibly rich source of information provided by language about a diverse range of phenomena, including sensory and quantitative relations such as color similarity and physical distances between locations in space (e.g., Marjieh et al., 2023).

The respective contributions of direct first-person experience and social learning to intuitive theory construction in humans is notoriously difficult to disentangle. In most cases, many types of evidence are simultaneously available and partly redundant. In the current study we approach this problem by comparing intuitive theories across adults with drastically different access to relevant first-person sensory experience, namely congenitally blind and sighted adults. Blind and sighted people share similar cognitive architecture, linguistic exposure, and cultural context, but blind people lack input from direct experience of seeing.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Comparing sighted and blind adults therefore offers a naturalistic test of how first-person sensory access contributes to qualitative and quantitative aspects of intuitive theories.

An intuitive theory of perception is one of the earliest developing aspects of 'theory of mind' and has even been hypothesized to provide a “stepping-stone” to understanding more abstract elements of the mind (e.g., beliefs; Gopnik et al., 1994; Meltzoff & Gopnik, 2013). Eighteen-month-old infants understand that looking is intentional and that someone else’s vision can be occluded by an opaque screen or blindfold but not by a transparent window (Brooks & Meltzoff, 2006; Butler et al., 2000). By 4.5 years old, children understand that an observer positioned closer to a small target would perceive it qualitatively better than someone situated further away (Flavell et al., 1980). However, previous research has not addressed the extent to which children or adults make quantitative inferences about perception: knowing not just that things are harder to see when they are further away, but by how much.

Intuitive theories of perception may be constructed by children at least in part based on first-person perceptual experience (Meltzoff & Brooks, 2008; Pillow & Flavell, 1986; Sommerville et al., 2005). For example, 12-month-old infants initially follow the gaze of a blindfolded adult but stop doing so after their own view has been blocked with a blindfold (Meltzoff & Brooks, 2008). Analogously, at three months of age, infants’ perception of other people’s actions is informed by their own action experience (Sommerville et al., 2005) . This evidence raises the question of what theories of perception would be like in the absence of first-person access: would they be qualitatively different, or perhaps differ only in their quantitative details?

In emphasizing the primacy of first-person experience, Empiricist philosophers, many 20th century educators, and contemporary psychologists have long believed that people who are blind have limited and fragmentary understanding of visual phenomena, such as colors, light, and visual perception (e.g., Berkeley, 1948; Hobbes, 1641, 1984; Hume, 1739, 1978; Locke and Nidditch, 2011). Contrary to this idea, seminal work by Landau and Gleitman (1985) showed that even young blind children appropriately use color adjectives and have nuanced understanding of ‘visual’ verbs such as “look” and “see” (see also Bedny et al., 2019; Kim et al., 2019; Shepard & Cooper, 1992). Blind adults also make subtle temporal distinctions between verbs like “glance” vs. “stare” in semantic similarity judgments (Bedny et al., 2019). However, most of the available research has only tested understanding of individual visual words (e.g., what is the meaning of “look”?) or visual facts (e.g, what color are bananas?). This leaves open the question of how far such knowledge extends in blindness. As mentioned above, sighted people’s understanding of vision goes far beyond such lexical or memory-based knowledge. Whether people born blind possess similar inferentially rich knowledge about vision remains to be tested. A handful of largely informal studies with small samples of blind children suggest that they may have a poorer understanding of visibility relative to sighted individuals (Bigelow, 1991a; Bigelow, 1991b).

In the current study, we go beyond examining lexical or memory-based knowledge about vision and compare generative and quantitative inferences about visual perception across congenitally blind and sighted groups. We collected data from two groups of sighted people, one matched in age and education to the blind group and the other as a sighted reference group. This allowed us to ask whether blind and sighted people are more different from each other than two groups of sighted people. Specifically, we asked congenitally blind and sighted adults to judge the likelihood that an observer would be able to see a particular feature of another person (e.g. the color of their eyes, whether they were wearing a hat, or had a pea-sized mole), when looking at that person from different distances and for varying durations. We developed a probabilistic model of intuitive visibility that is grounded in basic aspects of psychophysics, evaluated the model's ability to fit the judgments from each group, and compared its estimated parameters values for distance, duration, and size across blind and sighted participants.

Based on prior work, we hypothesized that blind and sighted people would use qualitatively similar models to predict visibility, however, these models may differ quantitatively. Specifically, we hypothesized that both groups would know that things are harder to see as they get further away and smaller and when people look for a shorter amount of time. But, we also predicted that first-person experience would lead to different weighting of these variables: e.g., blind and sighted people might disagree about how much visibility falls off with distance. Our hypothesis was based on the general observation that perception is excellent at conveying precise quantitative information, while language is more effective at communicating qualitative information (Borghi et al., 2019; Boyer, 1998; Landau & Jackendoff, 1993; Pinker, 2003). For example, we commonly say that something is “near”, “far”, “short”, or “long” without specifying the precise quantitative degree. When language is used to communicate precise quantity, it often relies on culturally invented formal systems (e.g., 20 feet). We therefore predicted that people who are blind and sighted have intuitive theories of visual perception with largely the same logical structure, but would differ in their detailed weighting of properties that impact visibility.

Methods

Participants

Twenty congenitally blind (44±16 y), 20 age and education matched sighted (45±15 y), and 20 reference sighted (20±1 y) Native English-speaking adults participated in the visibility rating task. A separate group of 21 sighted participants (19±1 y) provided subjective estimates of feature size. All blind participants had at most minimal light perception from birth, without visual experience of color, shape, or motion. The cause of vision loss was a pathology of

Table 1. Example visual scenarios and questions.

Example Scenarios	Example Questions
Next to: Sarah is a sighted adult. She was walking to the grocery store and walked past a brick rowhome that was right up against the sidewalk. A person was sitting on the front stoop of the rowhome. Sarah glanced/stared at this person as she walked by them and then kept going.	**Sample question:** "How likely is it that [*Sarah*] would know whether [*this person's eyes were light blue or dark brown*], based on what she saw?" **Sized features ('ground-truth' measurement):** 1. Is wearing a red or blue t-shirt (3475 cm^2) 2. Is wearing a red or blue quarter-sized pendant (4.62 cm^2) 3. Is wearing a hat or not (247.32 cm^2) 4. Has light blue or dark brown eyes (2.26 cm^2) 5. Has a mole the size of a pea on their cheek (0.69 cm^2) 6. Has a mole the size of a dime on their cheek (2.52 cm^2) 7. Has light blond or dark brown hair (247.32 cm^2)
Across the street: Steven, a sighed adult, was walking to the barber shop. On his way, he passed by a grocery store, which was on the other side of the street from him. A person was sitting on a bench in front of the store. Steven glanced/stared at this person as he walked by and then kept going.	
Other end of block: A sighted woman named Emma was walking to the craft store. All the way at the other end of the block, there was a tall telephone pole. A person was standing next to the pole. Emma glanced/stared at this person and then kept going.	

the eyes or of the optic nerve. The study was approved by the institutional review board of Johns Hopkins University, and all participants provided informed consent to participate in the experiment.

Procedure

Visibility Rating Task Participants listened to brief verbal scenarios (3-4 sentences long) in which an observer perceives another person by looking at them (Table 1). After each scenario, participants were asked to rate, on a scale of 1 ("definitely not") to 5 ("definitely yes"), how likely it is that the observer would have perceived specific features of the other person on the basis of what they saw.

We predicted that visibility ratings would be affected by the distance between the observer and the observed, the duration of looking, and feature size. Across scenarios, we manipulated the distance between the observer and the observed (3-levels), as well as the duration of looking (2-levels). Distance was manipulated by describing the person being observed as "right next to" (closest), "on the other side of the street from" (intermediate), or "all the way at the other end of the block from" (farthest) the observer. Duration of looking was varied by using the verb "glance" (brief) or "stare" (extended). See Table 1 for example scenarios. Features varied in size from "pea-sized mole on the cheek" to "t-shirt" (with 7 features in total, as in Table 1). We additionally asked about 7 qualitative features without clear sizes (e.g., the person's age and mood). These features were not amenable to objective ground-truth modeling and are not reported in the current study. (Two additional control questions were asked after each scenario: whether the person being observed has a British or American accent, and whether they have 1 or 2 siblings. The intended answer for the control questions was "definitely not".)

Participants heard 2 scenarios for each of the 6 distance-duration combinations, resulting in a total of 12 visual scenarios. Each participants heard all of the scenarios in one of two pseudorandom orders, with different conditions interleaved. The experimenter read the instructions and trials aloud. Participants' verbal ratings of visibility were audio recorded and transcribed. Participants also performed the same task on control auditory scenarios where a sighted or a blind listener perceives features of another person by hearing. The groups performed similarly on these scenarios (data not included here).

Estimating Distance (Objective and Subjective Measures) Ground-truth distances were estimated through consultation of online data ("next to": 0.61m; "across the street": 15.24m; "other end of the block": 79.25m). To verify that our main groups did not differ in subjective estimates of distance, we collected subjective distance estimates using a duration-of-walking task. Participants were asked to estimate how long it would take an average person to walk the length of a block or across the street in seconds or minutes. We used a duration measure because it is not known whether blind and sighted people have similar representations of distance units such as feet or meters. Blind and sighted participants both judged that it would take much longer to walk a block in seconds (blind: M=180.20, SD=95.96; matched sighted: M=141.25, SD=104.15; reference sighed: M=159.25, SD=104.10) than across the street (blind: M=49.73, SD=67.58; matched sighted: M=35.45, SD=35.80; reference sighed: M=30.68, SD=26.47; main effect of distance: $F(1,20)=106.58, p<.001$).

There were no significant between-group differences at either distance level (all *ps* >.05).

Estimating Size (Objective and Subjective Measures) We estimated the ground-truth size of features through manual measurement and consultation of online sources (e.g., coin specifications: https://www.usmint.gov/learn/coin-and-medal-programs/coin-specifications). To assess whether people's subjective size judgments generally correlate with these objective estimates, a separate group of sighted participants who did not participate in the main experiment were asked to provide estimates of the relative size of the 7 features. Participants were given an outline of a human figure on a piece of paper and asked to draw the contours of each feature (e.g., "eyes", "hat"). Features were drawn individually on separate pages in 1 of 2 pseudorandom orders. The drawings were traced on an electronic writing pad and measured with an online area calculator (https://www.sketchandcalc.com/). There was a very high correlation between log-transformed ground-truth and subjective measurements of feature size ($r(5)=.99$, $p<.001$).

Analysis and Results

Correlation of Ratings Across Groups

The average visibility ratings for each group, together with 95% bootstrap confidence intervals computed from participant means, are shown in Figure 1 for the stimulus dimensions distance, duration, and size separately. The two sighted groups assigned similar ratings along all three dimensions, while congenitally blind individuals generally rated visibility as more probable.

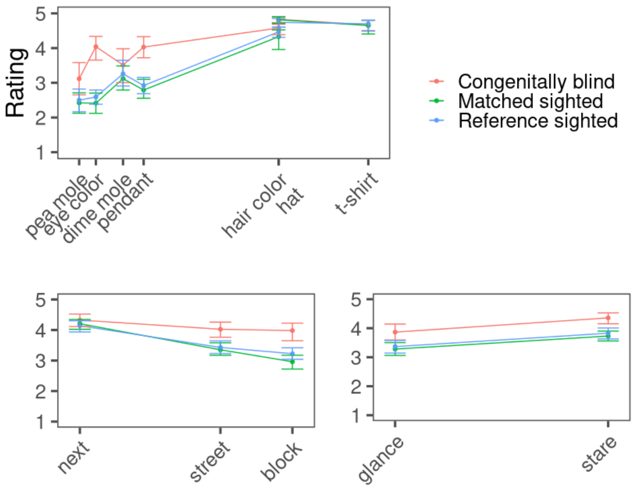

Figure 1: Mean group visibility ratings per size, distance, and duration levels.

Visibility ratings were highly correlated across the two groups of sighted participants. This was true of the group data in aggregate (Figure 2A), as well in a leave-one-subject out individual correlation analysis (results not shown here). Likewise, blind and sighted ratings were highly correlated with each other, although not quite as highly as sighted to sighted data (Figure 2A). All participants provided low ratings to the control questions (blind: M=1.05, SD=.27; matched sighted: M=1.06, SD=.32; reference sighted: M=1.05, SD=.33), which were excluded from subsequent analysis.

Modeling Visibility as a Function of Distance, Duration, and Size

We developed a probabilistic model of intuitive visibility, and a statistical link between this model and the rating task, to better understand the similarities and differences among the groups. Given the information provided by a scenario ($\mathbf{x}$), the model assigns a distribution on visibility (a latent scalar quantity) from the perspective of the observer, $p(v \mid \mathbf{x})$. Technically, the model assigns a normal distribution on log visibility, $\mathrm{Normal}(\log v \mid \log f(\mathbf{x}), \sigma)$, where $\log f(\mathbf{x})$ is the mean of the distribution, and the standard deviation is modeled here as scenario- and observer-independent.

The form of the function $f(\mathbf{x})$ is motivated by basic psychophysics. Under highly idealized conditions, the visibility of an object is determined by its size and distance to the observer through the well-known equation for visual angle (i.e., *2*atan[size/(2*distance)]*). When size is small relative to distance, so that the visual angle is also small, this equation can be closely approximated by the ratio *size/distance*. Our model of intuitive visibility uses the same multiplicative form, but computes the average visibility $f(\mathbf{x})$ from possibly many properties of scenario $\mathbf{x}$ and allows the exponent of each property to be fit to behavioral data (rather than constraining the exponents to +1 and –1):

$$f(\mathbf{x}) = \exp(\beta_0) \cdot \prod_{k=1}^{K} x_k^{\beta_k}$$

$$\log f(\mathbf{x}) = \beta_0 + \sum_{k=1}^{K} \beta_k \log(x_k)$$

The special case of the model employed here predicts visibility as a function of distance to the observer, duration of looking, and feature size. Distance and size were quantitative predictors with ground-truth values (as discussed earlier), logarithmically transformed and then centered. Duration was an effect-coded categorical predictor (-1 vs. +1). Other variables may well have affected visibility ratings in our experimental scenarios (e.g., asking about the color vs. presence of an object). The general form of the model could ultimately encompass these and many other properties known to impact visibility (e.g., lighting conditions and eyesight of the observer). To link the model to rating data, we assumed that rating probabilities were determined by ordinal probit regression (Bürkner & Vuorre, 2019; Kruschke, 2014) with a single set of fit thresholds applied to $\mathrm{Normal}(\log v \mid \log f(\mathbf{x}),$

σ) and σ set to 1. A hierarchical regression analysis was implemented in Stan (Carpenter et al., 2017) with group-level intercepts and predictor coefficients, participant-level intercepts, and weakly informative priors on all parameters. Inference was based on a single chain of 5,000 MCMC samples (1,000 warmup) from the posterior distribution specified by the analysis.

There were high group-level correlations between predicted and observed ratings in both sighted and blind groups (Figure 2B). All three groups had a negative coefficient for distance and positive group-level coefficients for duration and size (see Table 2). This supports a shared qualitative core of intuitive visibility regardless of first-person experience. The blind group differed from the two sighted groups in having a significantly larger intercept (i.e., a higher probability of visibility overall) and effects of size and distance that were significantly smaller in magnitude (i.e., with smaller rates of 'decay' as distance grows, or as size shrinks; all *p*s < .001). The groups did not differ here for the sparsely sampled predictor of duration.

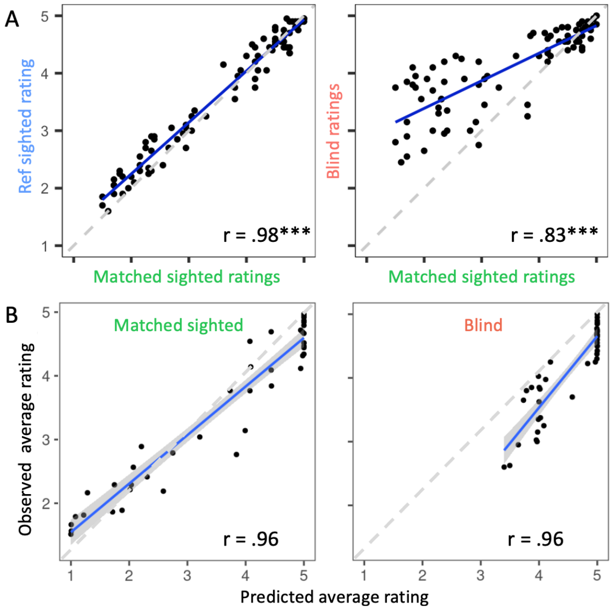

Figure 2: (A) Between-group correlations on observed ratings for each distance x duration x size x scenario combination, *** p<.001. (B) Group level correlations between predicted and observed ratings for each distance x duration x size combination.

We compared the results of the first regression analysis with two alternatives using the WAIC criterion as computed from the posterior samples. One alternative constrained the three groups to have the same effects of the predictors, allowing them to differ only in their intercepts. Relative to the first analysis (WAIC = 10279.5, SE = 110.0), this one fit the rating data much worse (WAIC = 10501.4, SE = 110.5). Another alternative included all possible group-level pairwise interaction terms (e.g., distance x size) in addition to the predictors above. This model provided a better fit (WAIC = 10168.7, SE = 112.4), suggesting that intuitive visibility is not a simple product of independent variables. Importantly, however, the group differences reported above remained in this more complex model.

Table 2. MCMC coefficients and 95% HPDI ranges.

Predictor	**Mean coefficient**	**95% HPDI interval**
Intercept		
Blind	1.42	[1.06 1.77]
Matched S	0.71	[0.42 1.02]
Ref S	0.81	[0.54 1.13]
Distance		
Blind	-0.12	[-0.15 -0.09]
Matched S	-0.36	[-0.39 -0.33]
Ref S	-0.26	[-0.29 -0.23]
Duration		
Blind	0.39	[0.33 0.45]
Matched S	0.33	[0.28 0.39]
Ref S	0.33	[0.28 0.39]
Size		
Blind	0.52	[0.47 0.56]
Matched S	0.83	[0.79 0.88]
Ref S	0.79	[0.74 0.83]

Discussion

We compared quantitative visibility inferences between congenitally blind and sighted adults. When listening to story-like scenarios about an observer who is "staring" or "glancing" at another person, blind and sighted people make highly consistent quantitative judgments about how likely the observer is to know properties of the person being observed (e.g., whether they have blue or brown eyes.) When blind and sighted people infer an observer's perceptual knowledge based on the observer's seeing experience, they rely on variables known to predict perceptibility in the real world: the distance of the observer from the observed ("right next to", "across the street", "a block away"), the duration of looking ("glance" vs. "stare"), and the size of the feature being viewed ("eye color" vs. "wearing a hat"). The direction of these effects is similar across groups, blind and sighted participants alike judged seeing as more likely when the observer was described as closer to the person being observed, when the observer looked for a longer duration, and when the observed feature was larger in size. However, people who are blind show different weighting of distance and size, as discussed in more detail below.

A simple probabilistic model based on a multiplicative (log-linear) combination of distance, duration, and size predicts sighted and blind human rating performance with a high degree of accuracy. Together, these results suggest that humans have a generative, quantitative model of how perception works and use it to make judgments about what someone is likely to know based on their visual experience: an intuitive theory of perception.

While here we included only three easily quantifiable variables in the model, other variables definitely affect visibility in general and could be incorporated in future work. For example, brightness contrast directly affects perceived visibility (Bozorg et al., 2016). When these variables are not specified, people may default to 'ideal' viewing conditions. Properties of the observer might also be incorporated. For example, sighted individuals preferentially allocate visual attention to parts of the body that convey social information, particularly the face (Hewig et al., 2008; Itier & Batty, 2009). We also did not distinguish between the relative visibility of detecting the presence of a feature (e.g., wearing a hat) and identifying its color (e.g., red/blue shirt). Whether and how blind and sighted people incorporate information about these and other variables into their visibility inferences is worth exploring.

Our results suggest that first-person sensory experience is not required to reason generatively about perception. This implies two separate conclusions. First, simulation using one's own visual experiences cannot be the only way to solve the visibly judgment problem. Sighted people could in principle solve the task by 'replaying' past visual experiences (Goldman, 2006; Gordon, 1995; Gallese & Goldman, 1998; Rizzolatti & Sinigaglia, 2008) but people born blind cannot and are unlikely to be using their visual system at all (e.g., (Sadato et al., 1996). Consistent with the idea that people use abstract theory-like representations to reason about perception, neuroimaging work suggests that people who are blind and sighted use a common neural mechanism to infer beliefs based on visual experiences (Bedny et al., 2009; Koster-Hale et al., 2014). Moreover, this mechanism is the same one that supports abstract reasoning about beliefs (Bedny et al., 2009). The current results go beyond this past work to show that the intuitive theory of perception used by people born blind is generative, quantitative, and detailed, even when it comes to vision-specific aspects of perception (e.g., quantitative effects of size). In future work it would be informative to test the neural basis of this type of reasoning. Meanwhile, the evidence suggests that humans have an abstract causal intuitive theory of perception used to reason about the sensory experiences of others. If this hypothesis is correct, congenitally blind and sighted people should also be able to provide causal explanations for visibility (e.g., explaining poor visibility as due to far distance), and to intervene to improve or hinder visibility (e.g., by moving the object closer to the viewer).

The present results also suggest that first-person experience is not necessary for acquiring the logical framework of an intuitive theory of perception. How do people born blind acquire this framework? Likely blind individuals use many sources of evidence, including analogy to other modalities. For example, we see and hear things less well when stimuli are further away. Such analogies are not as straightforward as they might first appear, however, since physical variables, such as distance and size, do not affect perception across modalities in the same ways (e.g., size has very different effects on auditory and tactile perception). For touch, distance cuts off where the arm ends, and size does not affect either tactile or auditory perceptibility in the same way that it affects vision.

We hypothesize that social learning from linguistic communication with sighted people plays a central role. Language is highly effective at communicating causal frameworks that are at the heart of intuitive theories (Pinker, 2003; Tooby & DeVore, 1987). A recent compelling demonstration of this phenomenon comes from large language models (LLMs), which though trained exclusively on text corpora, contain vast information about the world. Recent evidence demonstrates that these models contain information about visual (e.g., color) and auditory perception (e.g., pitch) (Abdou et al., 2021; Marjieh et al., 2023). Indeed, when presented with the task and scenarios from the current study, GPT4 shows human-like performance (Akshi et al., 2024), suggesting that in principle such learning is possible. However, GPT4 has access to vastly more linguistic evidence than any human, more memory and different learning. The acquisition of an intuitive theory of visual perception by people born blind is a compelling naturalistic demonstration of human learning from linguistic evidence. Together with the present results, this evidence suggests that social transmission via language effectively conveys the logical skeleton of an intuitive theory of perception in the absence of direct sensory access. Together with other evidence, this suggests the importance of language in intuitive theory transmission.

Language and perception, though partly redundant, are also complementary in terms of the information they convey. It has been suggested that language is less effective than perception at conveying precise, quantitative information (Borghi et al., 2019; Boyer, 1998; Landau & Jackendoff, 1993; Pinker, 2003). Spatial language frequently leaves out detailed representations of objects and their locations while offering schematic sketches of geometric relationships (Landau & Jackendoff, 1993). During communication people often augment their language with co-speech gesture, in part to make up for what language lacks (e.g., "I saw a fish this big") (Goldin-Meadow & Brentari, 2017). Consistent with these observations, we find that the quantitative weighting of distance and size is different across sighted and blind people. Relative to sighted people, people who are born blind tended to overestimate the informativeness of vision. This evidence suggests that sighted people use their own first-person sensory experience to 'tweak' the parameters of their intuitive theory of vision.

Overall, our results from visibility inferences in congenital blindness suggest that intuitive theories of perception are generative causal models that yield quantitative predictions. The logical structure of such a model can be acquired without direct access to the sensory modality specifically involved. Learning from language likely plays a role. First-person sensory experience, in turn, serves to fine-tune the weighing of the relevant variables.

Acknowledgments

We would like to acknowledge all of our participants, the National Federation of the Blind, and the blind community, for making this research possible. We are grateful to Shari Liu for her astute feedback on an earlier version of this draft, and Clarissa Alfonso and Gabriel Pernell for their assistance with data collection. This work was supported by the National Eye Institute (1R01EY033340; 1R01EY027352).

References

Abdou, M., Kulmizev, A., Hershcovich, D., Frank, S., Pavlick, E., & Søgaard, A. (2021). Can language models encode perceptual structure without grounding? A case study in color. *Proceedings of the 25th Conference on Computational Natural Language Learning*, (pp.109–132). Philadelphia, PA: Association for Computational Linguistics.

Akshi, Pernell, A., Wang, Z., & Bedny, M. (2024, July 24-27). *Language captures rich information about perceptibility: Evidence from LMMs and humans* [Poster Presentation]. Cognitive Science Society, Rotterdam, the Netherlands. https://cognitivesciencesociety.org/cogsci-2024/

Bigelow, A. E. (1991). Hiding in blind and sighted children. *Development and Psychopathology*, *3*(3), 301-310.

Bigelow, A. E. (1991). The effects of distance and intervening obstacles on visual inference in blind and sighted children. *International Journal of Behavioral Development*, *14*(3), 273-283.

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., & Saxe, R. (2019). There's more to "sparkle" than meets the eye: Knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition*, *189*, 105–115.

Bedny, M., Pascual-Leone, A., & Saxe, R. R. (2009). Growing up blind does not change the neural bases of Theory of Mind. *Proceedings of the National Academy of Sciences*, *106*(27), 11312–11317.

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, *29*, 120–153.

Boyer, P. (1998). Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist*, *100*(4), 876–889.

Bozorg, S. C., Maksimainen, M., Tetri, E., Kosonen, I., & Luttinen, T. (2016). The effects of dimmable road lighting: A comparison of measured and perceived visibility. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 141–156.

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology*, *38*(6), 958.

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101.

Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, *1*(4), 359–377.

Carey, S. (2009). *The origin of concepts*. Oxford, UK: Oxford University Press.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).

Flavell, J. H., Flavell, E. R., Green, F. L., & Wilcox, S. A. (1980). Young children's knowledge about visual perception: Effect of observer's distance from target on perceptual clarity of target. *Developmental Psychology*, *16*(1), 10–12.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493-501.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429.

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*, e46.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford, UK: Oxford University Press.

Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. In L. M. Antony & N. Hornstein (Eds.), *Chomsky and his critics*. Oxford, UK: Blackwell Publishing Ltd.

Gopnik, A., & Meltzoff, A. N. (1998). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gopnik, A., Slaughter, V., & Meltzoff, A. (1994). Changing your views: How understanding visual perception can lead to a new theory of the mind. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108.

Gordon, R. M. (1995). Simulation without introspection or inference from me to you. In M. Davies & T. Stone (Eds.), *Mental simulations.* Hoboken, New Jersey: Wiley-Blackwell.

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, *25*(10), 896–910.

Hewig, J., Cooper, S., Trippe, R. H., Hecht, H., Straube, T., & Miltner, W. H. R. (2008). Drive for thinness and attention toward specific body parts in a nonclinical sample. *Psychosomatic Medicine*, *70*(6), 729–736.

Hobbes, T. (1641/1984). Meditations on first philosophy, objections with replies. In R. Stootboff & D. Murdoch

(Eds.), *The Philosophical writings of Descartes.* Cambridge, UK: Cambridge University Press.

Hume, D. (1739/1978). *A treatise of human nature.* Oxford, UK: Oxford University Press.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843–863.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Kim, I., & Spelke, E. S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science*, *2*(3), 339–362.

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222.

Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: Perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition*, *133*(1), 65–78.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS and Stan* (2nd ed.). Cambridge, Massachusetts: Academic Press.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, Massachusetts: Harvard University Press.

Landau, B., & Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(2), 217–238.

Locke, J., & Nidditch, P. H. (2011). *An essay concerning human understanding.* Oxford, UK: Clarendon Press.

Marjieh, R., Sucholutsky, I., Rijn, P. V., Jacoby, N., & Griffiths, T. L. (2023). Large language models predict human sensory judgments across six modalities. ArXiv. https://doi.org/10.48550/arXiv.2302.01308

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: A training study in social cognition. *Developmental Psychology*, *44*(5), 1257–1265.

Meltzoff, A. N., & Gopnik, A. (2013). Learning about the mind from evidence: Children’s development of intuitive theories of perception and personality. In Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. V. (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (3rd ed.) Oxford, UK: Oxford University Press.

Pillow, B. H., & Flavell, J. H. (1986). Young children’s knowledge about visual perception: Projective size and shape. *Child Development*, 125–135.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution*. Oxford, UK: Oxford University Press.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions*. Oxford, UK: Oxford University Press.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526-528.

Shanon, B. (1976). Aristotelianism, Newtonianism and the physics of the layman. *Perception*, *5*(2), 241–243.

Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, *3*(2), 97–104.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants’ perception of others’ actions. *Cognition*, *96*(1), B1–B11.

Tooby, J. & DeVore, I. (1987) The reconstruction of hominid behavioral evolution through strategic modeling. In W.G. Kinzey (Ed.), *The evolution of human behavior: Primate models.* Albany, NY: SUNY Press.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, *21*(9), 649–665.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*(1), 337–375.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841–864.