**Why depth cue scaffolding?**

Humans employ a statistically optimal strategy for combining depth cues such as disparity and texture for horizontal (Knill & Sanders, 2003) and vertical (Hillis et al., 2004) slant discrimination. This is also true for the integration of visual and haptic information (Ernst & Banks, 2002).

Suggesting the presence of well-developed circuits for combining depth-cue information

As the conflict between two cues increases, the decision function changes from a linear model (Landy et al., 1995; Knill & Sanders, 2003; Hillis et al., 2004) to a Bayesian nonlinear cue integration model (Knill, 2007)

Humans employ a statistically optimal strategy for combining depth cues within vision (Knill & Sanders, 2003; Hillis et al., 2004) and across different modalities (Ernst & Banks, 2002, van Beers et al., 2011). When sensory noise and cue conflict is relatively low, the optimal model is a linear weighted average of each cue, which aims at reducing variance (Ernst & Banks, 2002; Knilll & Sanders, 2003; Hillis et al., 2004). However, optimal integration seems to be adaptable and can be influenced through feedback (Jacobs & Fine, 1999; Ernst et al., 2000; Atkins et al., 2001), though feedback is not a requirement (Knill, 2007b). This reweighting of context-specific priors seems to be highly adaptable (Jacobson & Fine, 1999; Knill, 2007a; Knill, 2007b) and can occur at a relatively fast rate (vanBeers et al., 2011). Importantly, by pairing haptic feedback with binocular cues, observers can be trained to rely on binocular disparity cues (Ernst et al., 2000; Knill, 2007b; Vedamurthy et al., 2016).

Fusion is an important computation performed by the dorsal visual stream

However, when cue conflict is large, the optimal model shifts to a Bayesian nonlinear model (Knill, 2007).

“Subjects can learn context-specific prior, which suggests that our visual system is flexible enough to apply different prior models to scenes in different context” (Knill, 2007a; Knill, 2007b)