Disentangling Spillover: Individual Differences Across Multiple Dimensions

Cui Ding[1], Ethan Gotlieb Wilcox[2], Lena A Jäger[1]

[1]University of Zurich, [2]Georgetown University

Background: Spillover effect happens when the processing difficulty of one word affects reading times on the following, but its causes and stability across individuals are still unknown. Classic reading models, such as E-Z Reader and SWIFT, link spillover to oculomotor planning, assuming it remains stable across individuals. Sentence processing models, e.g., surprisal theory, often treat it as noise. Some recent studies suggest that spillover may reflect deeper cognitive processes, but it is still unclear whether it reflects a stable reader trait, a temporary state, or an interaction of both. This project measures the stability of individual spillover effects; tests how they vary across different reading contexts, and whether they are connected to an individual's reading behaviors or psychometric profiles.

Method: We used six eye-tracking datasets that allow us to compare each participant's spillover effects across different reading situations. In InDiCo (German), participants read in two sessions, with a two-week interval between them. In PoTeC (German), participants read both within and outside their expertise domain. In OneStop (English), participants did both normal naturalistic reading and repeated reading. In GECO-NL and GECO-ZH, Dutch and Chinese participants read in both their native language (L1) and English (L2). In HKC (Chinese), participants read both single sentences and full paragraphs. For each participant, we estimated spillover effects of word length, lexical frequency, surprisal, and their stability using a Bayesian hierarchical model. We then tested how these spillovers related to skipping and regressions, and whether they could be predicted from psychometric profiles.

Results: Spillover effects showed different levels of stability across reading contexts **(Fig. 1)**. Word length spillover was the most stable, while frequency and surprisal spillover were less consistent. People showed different spillover patterns depending on the reading context – they did not adopt the same strategy in all situations. Spillover effects were more stable across sessions, domains, and when the L1 and L2 were similar (e.g., Dutch-English), but less stable across distant languages (e.g., Chinese-English), reading regimes, and context lengths. Individual spillover effects were also related to reading behavior **(Fig. 2)**. Readers with stronger word length spillover tended to skip fewer and regress more. Interestingly, readers with stronger surprisal spillover skipped more words and made fewer regressions. Among all measured cognitive profiles, reading fluency was the strongest predictor: more fluent readers are less sensitive to previous word length, but more sensitive to previous word's surprisal **(Fig. 3)**.

Discussion: Our results show that spillover effects are not fixed but vary across readers and contexts. Spillover is not just noise: its higher stability across sessions and domains suggests a stable reading trait, while lower stability across distant languages, formats, and tasks suggests it is context sensitive. The opposite patterns between word length versus surprisal spillover, both in reading behavior and cognitive predictors, suggest that they come from different underlying mechanisms. Word length spillover may reflect bottom-up processing, such as visual input or motor planning, while surprisal spillover is linked to higher-level prediction and integration. Overall, spillover reflects both a person's general reading strategy and their adaptation to the reading context.

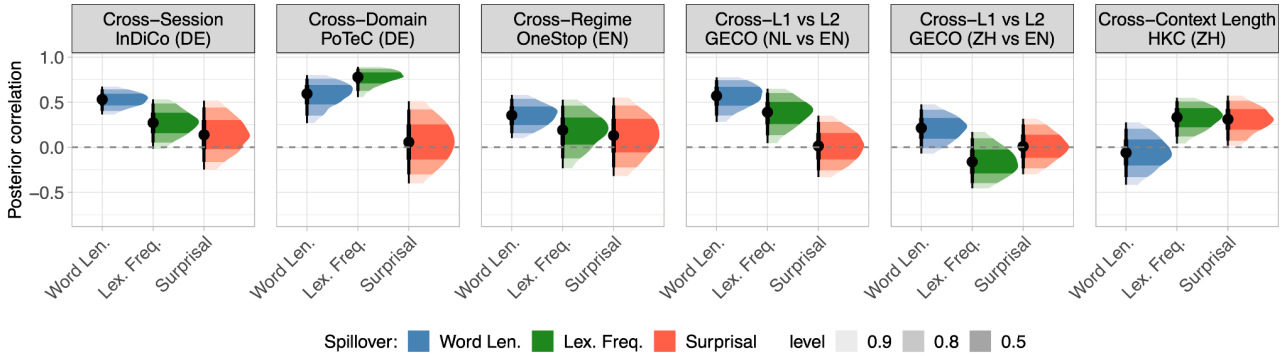

Fig. 1: Measurement reliability of individual spillover effects: posterior distributions of cross-context correlation coefficients for word length, lexical frequency, and surprisal effects on gaze duration. Reliability is assessed across sessions (InDiCo), expertise domains (PoTeC), reading regimes (OneStop), L1 vs L2 (GECO-NL, GECO-ZH), and context length (HKC).

Fig. 2: Relationship between individual spillover effects and reading patterns. Each panel shows the correlation between individual spillover effect sizes (top: word length; bottom: surprisal) and individual reading behaviors – skip rate (top row) and regression rate (bottom row) – across datasets. Different colors represent different contexts within each dataset (e.g., session 1 vs. session 2 in InDiCo; sentence vs. paragraph reading in HKC). Pearson's r and significance levels are annotated.
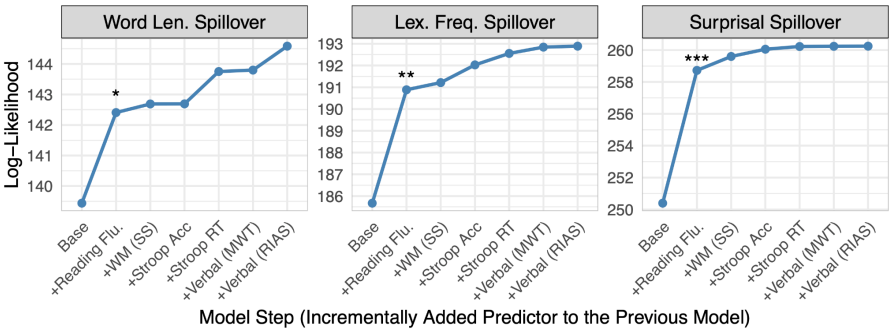
Fig. 3. Predicting individual spillover effects from psychometric profiles using IndiCo. Log-likelihood increases as reading fluency, working memory, cognitive control, and verbal intelligence are added to models of spillover (word length, frequency, surprisal). "*" mark significant predictors in the full model.