Web Supplementary Material of the article

"An application of procedural knowledge space theory to the study of strategic planning in board games"

Simulation Study

The simulation study presented in this section aims to assess the ability of the MSPM framework to produce unbiased parameter recovery in the context of a two-player game, specifically tic-tac-toe. A secondary goal is to evaluate the capacity of model selection criteria to discriminate between two knowledge spaces obtained from abstract representation of the same original problem space. The results shown in this section are the foundation for the interpterion of the empirical application.

Simulation design and data set generation

In the simulation study, data were generated under six conditions resulting from the combination of three knowledge spaces $\mathcal{K}_1^a$, $\mathcal{K}_2^a$ and $\mathcal{K}_b$, and two error levels (low and high). Table 1 displays the six conditions.

The error level in the data was manipulated through the transition probabilities β_{ij} and η_{ij} of the model. Specifically, for each $i \in S \backslash \{f, g\}$, the β_{if} parameters were drawn at random from a uniform distribution with intervals $(0, .1]$ and $(0, .3]$ depending on the conditions (low and high). Similarly, the η_{if} parameters were randomly drawn from a uniform distribution with intervals $[.7, 1)$ and $[.9, 1)$. Different ranges were chosen for β_{if} and η_{if} because they have distinct interpretations: β_{if} represents careless errors

Table 1: Design of the simulation study used for generating the data. Column one displays the condition number, column two displays the knowledge space, columns three and four show the error interval used for generating the data.

Condition number	$\mathcal{K}_t$	β_{if}	η_{if}
1	$\mathcal{K}_1^a$	(0,.1]	[.9,1)
2	$\mathcal{K}_1^a$	(0,.3]	[.7,1)
3	$\mathcal{K}_2^a$	(0,.1]	[.9,1)
4	$\mathcal{K}_2^a$	(0,.3]	[.7,1)
5	$\mathcal{K}^b$	(0,.1]	[.9,1)
6	$\mathcal{K}^b$	(0,.3]	[.7,1)

and is therefore expected to be small, while η_{if} represents non-lucky guesses and is thus expected to be close to one. The remaining β_{ij} and η_{ij} were drawn at random from a uniform distribution and then normalized to sum up to $1 - \beta_{if}$ and $1 - \eta_{if}$, respectively. For these reasons, conditions where $\beta_{if} \in (0, .1]$ and $\eta_{if} \in [.9, 1)$ represent a low error scenario, whereas conditions where $\beta_{if} \in (0, .3]$ and $\eta_{if} \in [.7, 1)$ represent a high error scenario. By construction, each pair (β_{ij}, η_{ij}) with $j \neq f$ also satisfies the inequality $\beta_{ij} - \eta_{ij} \geq 0$.

In each condition, a knowledge space $\mathcal{K}_t \in \{\mathcal{K}_1^a, \mathcal{K}_2^a, \mathcal{K}_b\}$ was used to generate the data. A knowledge state K was first randomly selected from $\mathcal{K}_t$. For each item $(s, g) \in Q$, a solution path was then generated by sequentially applying transition probabilities. If $(s, g) \in K$, transitions are governed by the probabilities β_{ij}; otherwise, they are governed by the probabilities η_{ij}. Each path terminated upon reaching either the goal state g or the failure state f. This procedure follows the data simulation algorithm

described in (Stefanutti, de Chiusole, & Brancaccio, 2021).

For each of the six conditions, 500 datasets were generated, yielding a total of $6 \times 500 = 3000$ simulated data sets.

Methods

Model identifiability was assessed using an non-formal two-step procedure. First, each model was re-estimated from 50 different sets of initial values, yielding 50 solutions with equal likelihood. Second, the standard deviations of the parameter estimates across these solutions were computed. When parameters are identifiable, these standard deviations should be zero up to numerical round-off. This approach is often used when a formal test of (local) identifiability is not available (see e.g, Stefanutti, de Chiusole, Anselmi, & Spoto, 2020; Stefanutti et al., 2021).

The goodness of parameter recovery for each knowledge structure was evaluated only under conditions where the knowledge structure used for model fitting matched the one used to generate the data. In these cases, the bias and variance of the parameter estimates were computed across 500 simulated datasets. The estimated bias for each β_{ij} was computed as follows:

$$bias(\hat{\beta}_{ij}) = \bar{\beta}_{ij} - \beta_{ij}, \tag{1}$$

where β_{ij} denotes the parameter value used for data generation and $\bar{\beta}_{ij}$ is the arithmetic mean of the 500 parameter estimates $\hat{\beta}_{ij}$. The standard deviation of $\hat{\beta}_{ij}$ was computed as

$$\sigma(\hat{\beta}_{ij}) = \sqrt{\frac{1}{500}\sum_{k=1}^{500}(\hat{\beta}_{ij}^{(k)} - \bar{\beta}_{ij})^2}, \tag{2}$$

where $\hat{\beta}^{(k)}ij$ is the estimate of the transition probability obtained from the k-th simulated dataset. Note that $\sigma(\hat{\beta}_{ij})$ represents also the standard deviation of Bias$(\hat{\beta}_{ij})$. The bias and standard deviation of $\hat{\eta}_{ij}$ were computed analogously.

In conditions 1–4, the MSPM was fitted using both $\mathcal{K}_1^a$ and $\mathcal{K}_2^a$. To assess model selection performance, the Akaike Information Criterion (AIC; Akaike, 1973) and the corrected Akaike Information Criterion (AICc; Hurvich & Tsai, 1989) were computed for the two competing models. The selection rate p_{AIC} is defined as the proportion of 1000 replications in which the AIC favored the data-generating knowledge space $\mathcal{K}_t$. Similarly, p_{AICc} is defined as the proportion of 1000 replications in which the AICc favored the data-generating knowledge space. For example, in condition 1, $p_{AIC} = .80$ would indicate that the model based on $\mathcal{K}_1^a$ is selected in 800 out of 1000 data sets, whereas in condition 3 the same value would indicate that the model based on $\mathcal{K}_2^a$ is selected in 800 out of 1000 data sets.

Results

The simulation results revealed a single instance of possible local non-identifiability in the knowledge space $\mathcal{K}_1^a$, where the parameter $\eta_{22,45}$ exhibited a standard deviation of .11. All remaining parameters within $\mathcal{K}_1^a$, as well as those estimated for the other two knowledge spaces, were identifiable, with standard deviations less than or equal to .02.

Figure 1 displays the bias of the parameter estimates under low-error (conditions 1, 3, and 5) and high-error (conditions 2, 4, and 6) conditions, shown in blue and red, respectively. The left panels correspond to the β_{ij} transition probabilities, and the right panels to the η_{ij} transition probabilities. The uppers panels report results for conditions 1 and 2, the central panels for conditions 3 and 4, and the bottom panels for

conditions 5 and 6. In each panel, the vertical axis lists the pairs (i, j) corresponding to the parameter estimates $\hat{\beta}_{ij}$ or $\hat{\eta}_{ij}$, with $i \in S$ and $j \in E(i)$, while the horizontal axis reports the associated bias. For clarity, only free transition probabilities are displayed (e.g., if the constraint $\hat{\eta}_{ij} + \hat{\eta}_{if} = 1$ holds, only $\hat{\eta}_{ij}$ is shown).

It should be noted that most parameters are accurately recovered, with bias values equal to or close to zero. In condition 2, the $\eta_{i,g}$ parameters showed slightly larger standard deviations and small positive biases. This result is likely due to the low probability of observing incorrect responses for the simplest problem in the structure (i.e., the problem involving a direct transition to the goal state). In addition, in condition 2, five $\beta_{i,j}$ parameters exhibited small negative biases, namely $\beta_{12,8}$, $\beta_{13,10}$, $\beta_{15,30}$, $\beta_{25,45}$, and $\beta_{34,45}$. However, with the exception of $\beta_{25,45}$, these biases tended to disappear in condition 1, which is a low error condition. The only other notable bias is observed in condition 4 for the parameter $\beta_{12,8}$. We don't have any explanation to offer with respects to these biases, however some conjecture can be provided. For instance, for every simulated sample the parameters were estimated only once. This may increase the risk to end up on a local maximum. In this case the parameters estimated would be biased for some of the parameteres.

In all conditions, the standard deviation of the $\beta_{i,j}$ parameter estimates tended to be larger for parameters farther from the goal (i.e., those based on correct responses to more difficult problems) and smaller for parameters closer to the goal (i.e., those based on easier problems). The opposite pattern is observed for the $\eta_{i,j}$ parameters, where the standard deviation of the bias was smaller for parameters farther from the goal and larger for those closer to it.

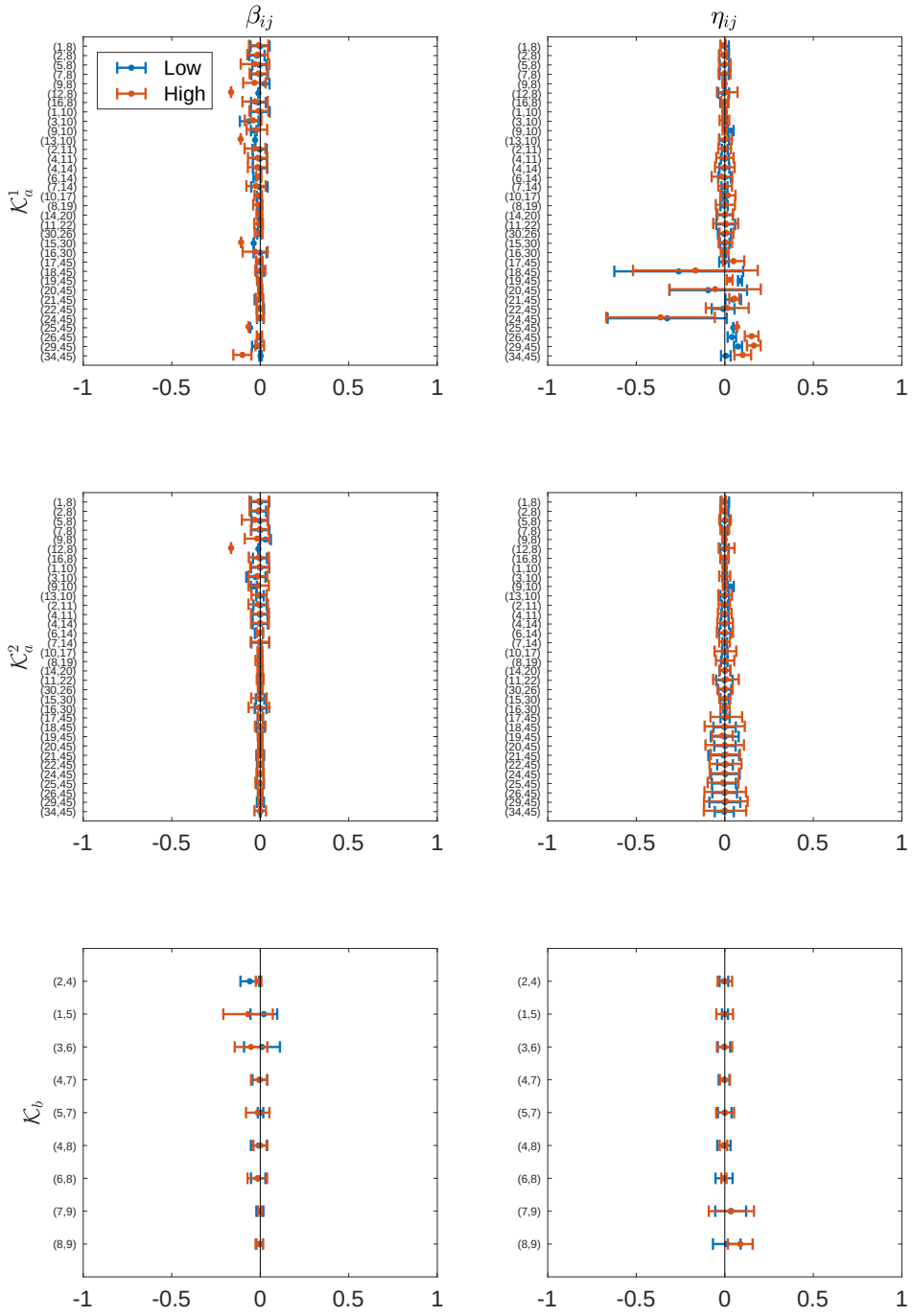


Figure 1: Bias of the MSPM parameter estimates under low- and high-error conditions, shown in blue and red, respectively. Left panels display the β_{ij} transition probabilities, and right panels the η_{ij} transition probabilities. The upper panels correspond to conditions 1 and 2, the central panels to conditions 3 and 4, and the bottom panels to conditions 5 and 6. The vertical axis reports the pairs (i, j) representing the estimated parameters $\hat{\beta}_{ij}$ or $\hat{\eta}_{ij}$, with $i \in S$ and $j \in E(i)$, while the horizontal axis shows the bias. Only free transition probabilities are displayed.

Table 2 reports the selection rates p_{AIC} and p_{AICc} for conditions 1–4. The p_{AIC} values (column 1) consistently favored the smaller model (i.e., knowledge space $\mathcal{K}_a^2$), regardless of the knowledge space used to generate the data, making it an ineffective selection criterion. In contrast, p_{AICc} consistently identified the data-generating knowledge space across all four conditions, indicating that it represents a more reliable criterion for model selection in this context.

Table 2: Selection rates for conditions 1–4.

Condition	p_{AIC}	p_{AICc}
1	0.00	1.00
2	0.00	1.00
3	1.00	1.00
4	1.00	1.00

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Academiai Kiado.

Hurvich, C. M., & Tsai, C. (1989, 06). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307. doi: 10.1093/biomet/76.2.297

Stefanutti, L., de Chiusole, D., Anselmi, P., & Spoto, A. (2020). Extending the basic local independence model to polytomous data. *psychometrika*, *85*(3), 684–715. doi: https://doi.org/10.1007/s11336-020-09722-5.

Stefanutti, L., de Chiusole, D., & Brancaccio, A. (2021). Markov solution processes: Modeling human problem solving with procedural knowledge space theory. *Journal of Mathematical Psychology*, *103*, 102552. doi: https://doi.org/10.1016/j.jmp.2021.102552