

MathPsych / ICCM 2024
mathpsych.org
July 2024

Predicting complex problem solving performance in the tailorshop scenario

Marco Ragni ~ TU Chemnitz

- ▶ Complex systems are present in our everyday lives
 - ▶ Societal, economic, environmental ...
- ▶ Complex Problem Solving (CPS) behavior studied with Dynamic Decision-Making (DDM) tasks
- ▶ Computer simulations - *Microworlds*
 - ▶ More realistic environments
 - ▶ More complex dependencies between variables

- ▶ Microworld simulating a tailorshop [1-5]
- ▶ Role of a tailorshop manager for 12 months
 - ▶ Purchasing raw materials, managing production capacity, maximizing profit by selling shirts
- ▶ 24 interconnected variables - 21 visible to participants, 12 directly manipulable
- ▶ Used to explore problem-solving processes, intelligence and professional performance
- ▶ Success defined as consistent increase in company value over months (first month excluded)

Tailorshop

Month 2 of 12

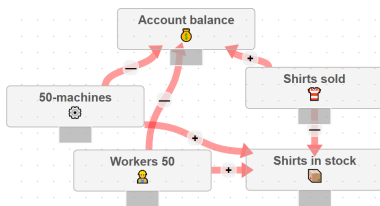
Indicator	Value	Planning	Action	Info
Account balance 🏠	169528 ↑			ⓘ
Shirts sold 🟢	439 ↑			ⓘ
Shirts in stock 🟡	64 ↓			ⓘ
Shirt price	52		- +	ⓘ
Sales outlets 🏢	1		- +	ⓘ
Outlet location 🗺️	City		- +	ⓘ
Workers 50 👤	10 ↑		- +	ⓘ
Workers 100 👤	0		- +	ⓘ
Salary 💰	1080		- +	ⓘ
Social costs / worker 📄	50		- +	ⓘ
Worker satisfaction % 😊	59 ↑			ⓘ

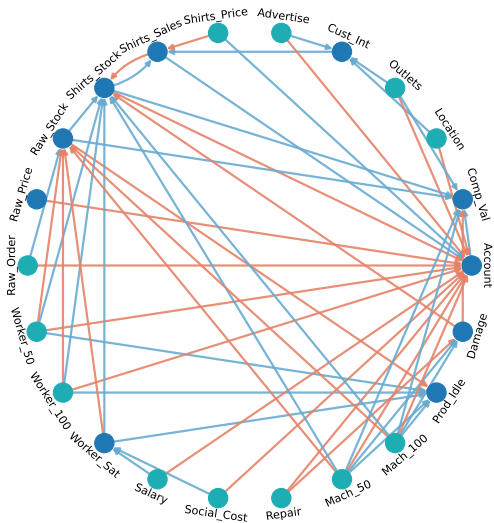
Indicator	Value	Planning	Action	Info
Company value 📄	269785 ↑			ⓘ
Customer interest ❤️	703 ↓			ⓘ
Raw materials in stock 🏢	244 ↑			ⓘ
Raw material price	8 ↑			ⓘ
Raw materials order 📄	650 ↑		- +	ⓘ
Advertising expenses 📄	2800		- +	ⓘ
50-machines ⚙️	10		- +	ⓘ
100-machines 🏭	0		- +	ⓘ
Repair & Service 🔧	1200		- +	ⓘ
Machine damage % ⚡	12 ↑			ⓘ
Production downtime % ⚠️	0			ⓘ

Continue

- ▶ “One-item-testing” of one large, complicated scenario is not good for CPS research [6]
 - ▶ More difficult to detect individual differences
- ▶ Predictive Modeling Perspective:
 - ▶ How does prior knowledge and individual characteristics influence behavior?
 - ▶ Are there any action patterns that can serve as a base for modeling endeavours?
 - ▶ Is participants’ performance predictable and how suitable is the Tailorshop for predictive modeling of CPS in general?

- ▶ 52 students at TUC
- ▶ Participants were asked to draw a causal-map before and after the TS simulation
- ▶ TS simulation had a 6-month (rounds) exploration phase and a 12-month test phase
- ▶ Participants completed Need for Cognition (NFC) and Cognitive Reflection Task (CRT) [7, 8]
- ▶ They were also asked which variables they deem important after finishing TS

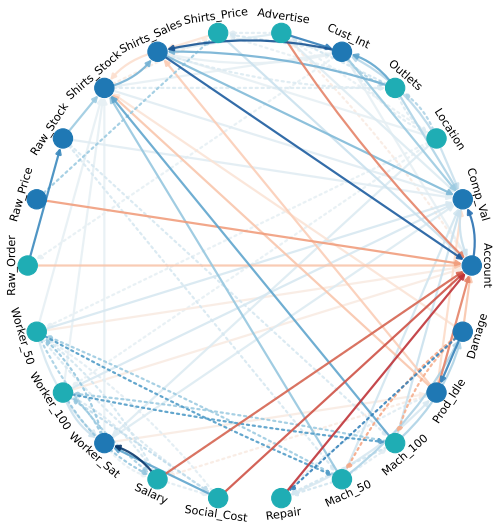




- Controllable Variables
- Derived Variables

Relationship Strength

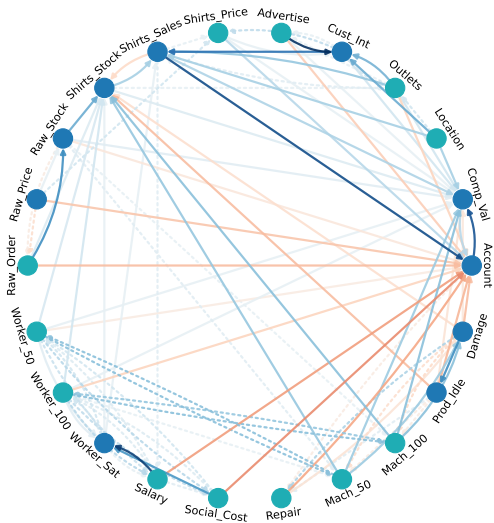




- Controllable Variables
- Derived Variables

Relationship Strength





- Controllable Variables
- Derived Variables

Relationship Strength



- ▶ Similarity of causal map to the actual simulation could indicate “correct” understanding
- ▶ Did participants adjust their assumptions after experiencing the tailorshop simulation?
- ▶ Node similarity calculation: Cosine similarities between adjacency vectors
- ▶ Not significant: Before = .247; After = .255 ($p = 0.781$)

- ▶ Importance of a variable w.r.t. company value
- ▶ # occurrences in all (cycle-free) paths leading to company value
- ▶ Average relevance reported by participants

	Var. Importance			Rel.
	Before	After	TS	
Company Value	(40.58)	(38.35)	(111)	-
Bank Account	24.19	25.31	67	-
Customer Interest	9.06	8.69	9	4.27
Shirts Sales	15.58	15.62	36	4.69
Shirts in Stock	7.19	9.77	72	3.79
Raw Material Price	0.46	1.46	0	3.56
Raw Material Stock	0.96	4.23	32	3.98
Worker Satisfaction	11.0	7.15	14	2.85
Production Idle	2.9	3.83	0	3.30
Damage	3.33	3.1	12	3.25

- ▶ Does a more “correct” understanding lead to better performance?
- ▶ Tailorshop performance was measured using the total difference in company value after 11 months
- ▶ One-sided Spearman correlation between similarity of Before graph to tailorshop graph showed a significant moderate correlation ($r = .264$; $p = .035$)
- ▶ Causal map correctness correlated with success → Potential for a predictive model?

- ▶ Support vector regression (SVR) was trained to predict the tailorshop performance based on the knowledge graph
- ▶ SVR was fitted using a leave-one-out cross-validation and had to predict individual results
- ▶ Knowledge graph was represented as an adjacency matrix

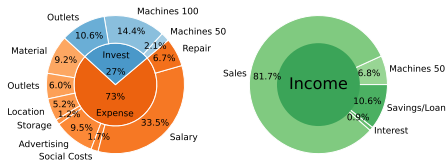
Predictor	MAE	RMSE	R^2
Performance Mean	0.298	0.378	0
Performance Median	0.295	0.381	-0.018
SVR (<i>Before graph</i>)	0.293	0.379	-0.007

- ▶ Model at baseline level, not suited to predict tailorshop performance
- ▶ Including individual traits (NFC, CRT) did not improve model performance

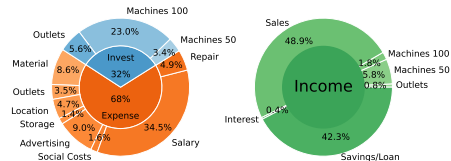
- ▶ Causal map was not sufficient to predict success in the tailorshop simulation
 - ▶ Is the causal map too limited?
 - ▶ Is the tailorshop too dynamic and complex to be predicted?
- ▶ We analyzed the relation between action patterns and success

- ▶ 79.17% were unprofitable (31.25% ended in debt)
- ▶ Only 20.83% were profitable
- ▶ What is the difference between them?

Profitable

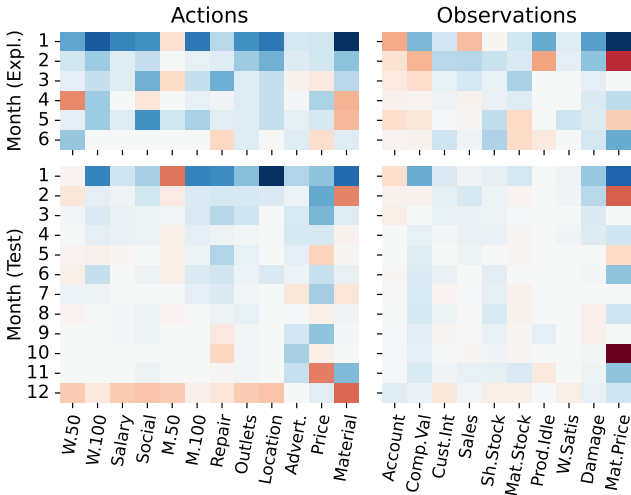


Unprofitable

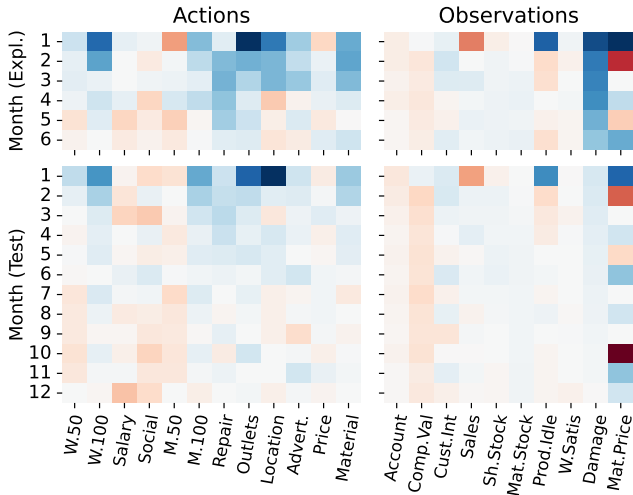


- ▶ Profitable TS rarely needed loans or savings
- ▶ Overall, investments and expenses are rather similar
- ▶ Problem seems to be finding the right moment in time

Profitable



Unprofitable



▶ Simple heuristic strategies as metrics:

1. *Upgrade machines*

→ Buy better machines, hire respective workers, and sell old machines

$$strategy1 = sign(\Delta M100 + \Delta W100) * sign(-\Delta M50)$$

2. *Avoid production loss*

→ buy raw material and invest in repair/maintenance

$$strategy2 = Material + Repair$$

▶ Both metrics correlate significantly with performance
(S1: $r = .310, p = .016$; S2: $r = .656, p < .001$)

▶ The simulation heavily depends on the first month

- ▶ Similarly to before: SVR used to predict tailorshop performance
- ▶ This time - based on the first month actions

Predictor	MAE	RMSE	R^2
Performance Mean	0.298	0.378	0
Performance Median	0.295	0.381	-0.018
SVR (<i>Before</i> graph)	0.293	0.379	-0.007
SVR (First Month Actions)	0.255	0.328	0.247

- ▶ The SVR now outperforms the baseline models with a positive coefficient of determination

- ▶ Causal Map was not sufficient
 - ▶ Participants did not seem to update their knowledge
 - ▶ Causal Maps did not allow to predict TS performance
- ▶ Tailorshop performance depends a lot on decisions in the first month
 - ▶ Can make other factors irrelevant → problematic for modeling
 - ▶ Reduces meaning of the actual management task
 - ▶ Even simple strategies and models are successful

- ▶ Complexity and dynamic environment makes the tailorshop prone to snowball effects
- ▶ Despite having several intermediate steps, the general state is remarkably determined by the initial actions
- ▶ Although CPS is important for cognitive modeling, the tailorshop simulation seems not to be well suited
- ▶ Needed: easily repeatable or only weakly self-reinforcing tasks

1. Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [The relation between test intelligence and problem solving success]. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie*, 189(1), 79-100.
2. Putz-Osterloh, W. (1983). Über Determinanten komplexer Problemlöseleistungen und Möglichkeiten zu ihrer Erfassung [On factors for complex problem solving and possibilities of their diagnosis]. *Sprache & Kognition*, 2, 100–116.
3. Funke, J. (1988). Using simulation to study complex problem solving: A review of studies in the FRG. *Simulation & Games*, 19(3), 277–303.
4. Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring performance in dynamic decision making. *Journal of individual differences*, 32, 225-233.

5. Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015). Sometimes less is more: Comparing the validity of complex problem solving measures. *Intelligence*, 50, 100-113.
6. Greiff, S., & Funke, J. (2009). *Measuring complex problem solving: The microdyn approach*. Office for Official Publications of the European Communities
7. Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2015). Deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition NFC-K [German short scale for measuring the construct Need for Cognition NFCK]. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*.
8. Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, 20(2), 147-168.