Reinforcing vs. informative feedback while controlling a dynamic system

Danilo Fum and Andrea Stocco

Laboratorio di Sistemi Cognitivi
Dipartimento di Psicologia
Università di Trieste

or:

something old, something new,
something borrowed, (and something blue)
about Sugar Factory,
and related issues.

Overview

- Something old:
  Sugar Factory's SOAR
- Something new:
  Fresh implications from a glorious model
  Original data from novel experiments
- Something borrowed:
  … let's a hundred flowers bloom
- Something blue:
  …

SF: State Of the Art Report

- Sugar Factory: an old, but still interesting (and sometimes surprising) research paradigm.

- People have to keep the production *P* of a simulated sugar factory on a target value by allocating an appropriate number of workers *W* to the job

- Discrete number of states [1..12] for both *P* and *W*, and discrete computational steps

- The system dynamics is controlled by the relation

  $$P_t = 2W_t - P_{t-1} + \varepsilon$$

- The task is made difficult by the existence of random noise ε, uniformly distributed with values {-1, 0, +1}.

SF: State Of the Art Report

- SF’s typical phenomenon: people progressively learn to control the system, but nobody seems to understand anything
- Initially assumed as a case for the existence of a separate implicit learning system
- Some have tried to explain the phenomenon by assuming that people rely on memorized records (instances) of their interactions with the system.

SF: State Of the Art Report

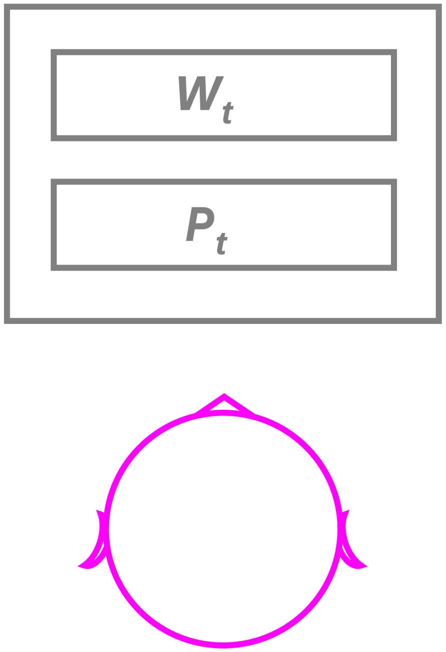
- We developed a procedural model of the SF task based on the ACT-R subsymbolic learning mechanism

A new model

Six productions compete according to a pure ACT-R learning scheme:

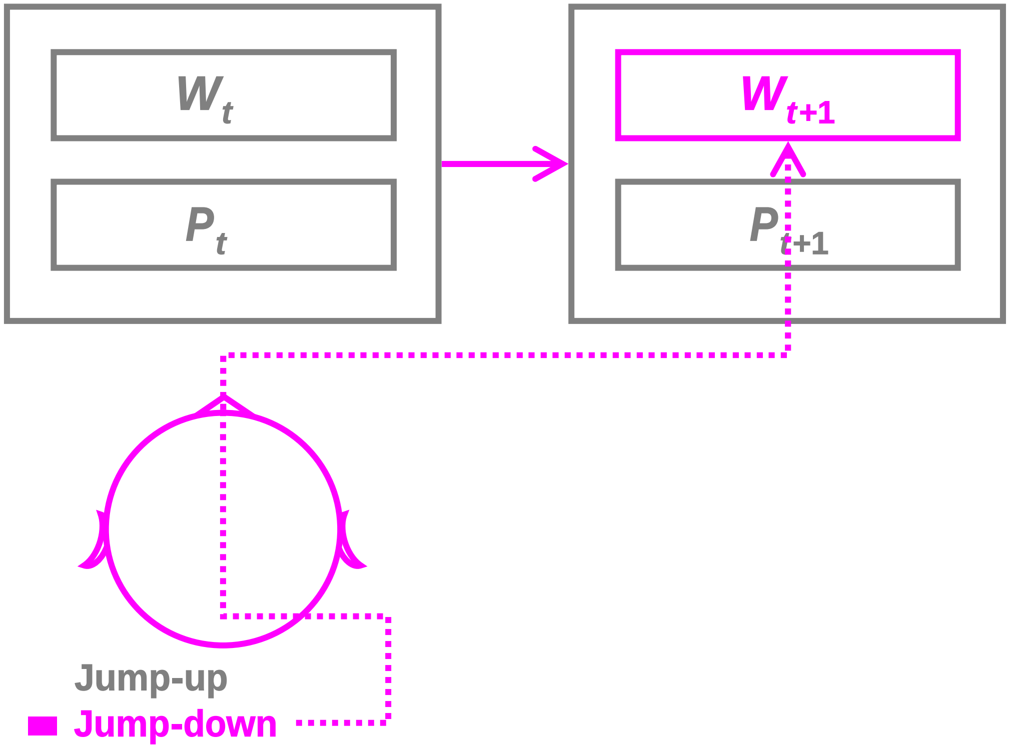
- choose-random: choose a random value between 1 and 12
- repeat-choice: repeat the previous *W* value
- stay-on-hit: if you hit the target, keep the same *W* value
- pivot-around-target: choose as *W* the value of the target (plus noise)
- jump-up: if your production *P* is below the target increase the value of *W*
- jump-down: if your production *P* is above the target decrease the value of *W*.

Interaction (at the beginning)



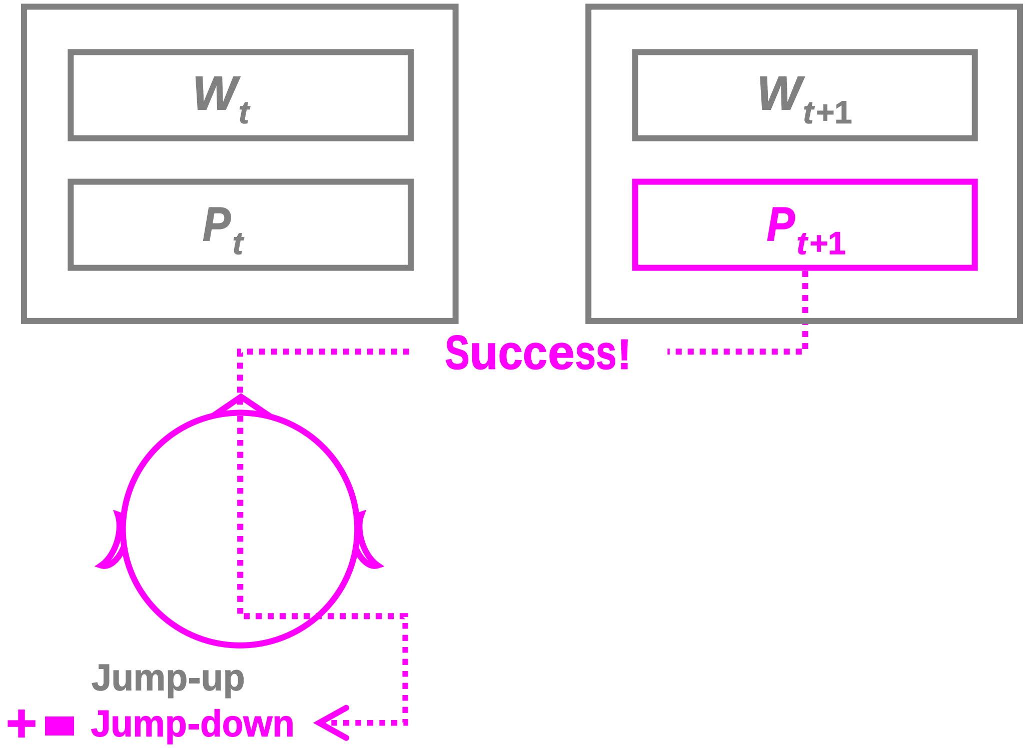
- Jump-up
- Jump-down
- Stay-on-hit
- Pivot-around-target
- Choose-random
- Repeat-choice

Interaction (at the beginning)



- Jump-up
- ■ Jump-down
- Stay-on-hit
- Pivot-around-target
- Choose-random
- Repeat-choice

Interaction (reinforcing feedback)



W_t

P_t

W_{t+1}

P_{t+1}

Success!

Jump-up

+ ▬ Jump-down

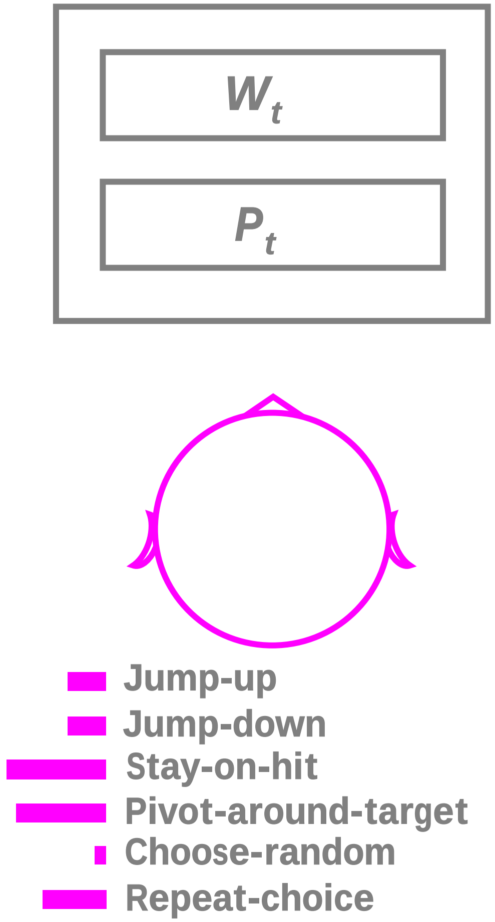
Stay-on-hit

Pivot-around-target

Choose-random

Repeat-choice

Interaction (after a while)



W_t

P_t

Jump-up

Jump-down

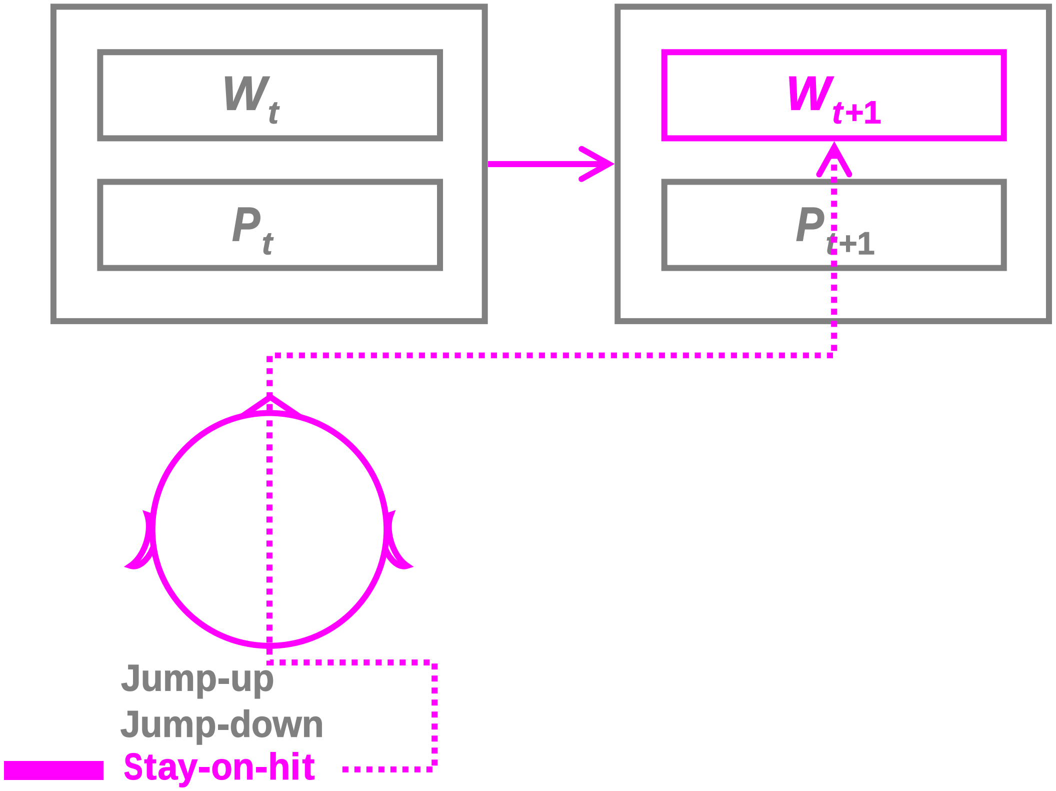
Stay-on-hit

Pivot-around-target

Choose-random

Repeat-choice

Interaction (next choice)



Jump-up
Jump-down
■ **Stay-on-hit**
Pivot-around-target
Choose-random
Repeat-choice

SF: State Of the Art Report

- We developed a procedural model of the SF task based on the ACT-R subsymbolic learning mechanism

- The model was able to replicate previous findings, and to explain three new effects concerning:
  - the role of different target values
  - the role of a change in the target value
  - the role of the evaluation criterion

SF: State Of the Art Report

- We developed a procedural model of the SF task based on the ACT-R subsymbolic learning mechanism
- The model was able to replicate previous findings, and to explain three new effects concerning:
  - the role of different target values
  - the role of a change in the target value
  - the role of the evaluation criterion
- **The model is congruent with a series of effects found in the literature that challenge the instance based models.**

Some critical predictions

- Playing the science game, we stretched the limits of the model, and we came up with some critical predictions concerning it, and the ACT-R architecture it is based upon

Some critical predictions

- Playing the science game, we stretched the limits of the model, and we came up with some critical predictions concerning it, and the ACT-R architecture it is based upon

- Basic ACT-R assumption concerning subsymbolic procedural learning mechanism: successful applications of a production increase its expected utility

Some critical predictions

- Playing the science game, we stretched the limits of the model, and we came up with some critical predictions concerning it, and the ACT-R architecture it is based upon

- Basic ACT-R assumption concerning subsymbolic procedural learning mechanism: successful applications of a production increase its expected utility

- **What would happen if we do not allow a production to experience success?**

A first scenario

- Let us suppose that, in setting up the SF task, we cheat and never allow participants to reach the target
- In this case, the productions used to execute the task should have no reason (and no occasion) to increase their expected utility, and the choice between them will be performed randomly
- As a result, no learning between the first and second phase should be expected.

An experiment

- We set up an experiment comparing two conditions:
  (a) the standard 9000-tons target vs.
  (b) a “snake” 9000-tons condition

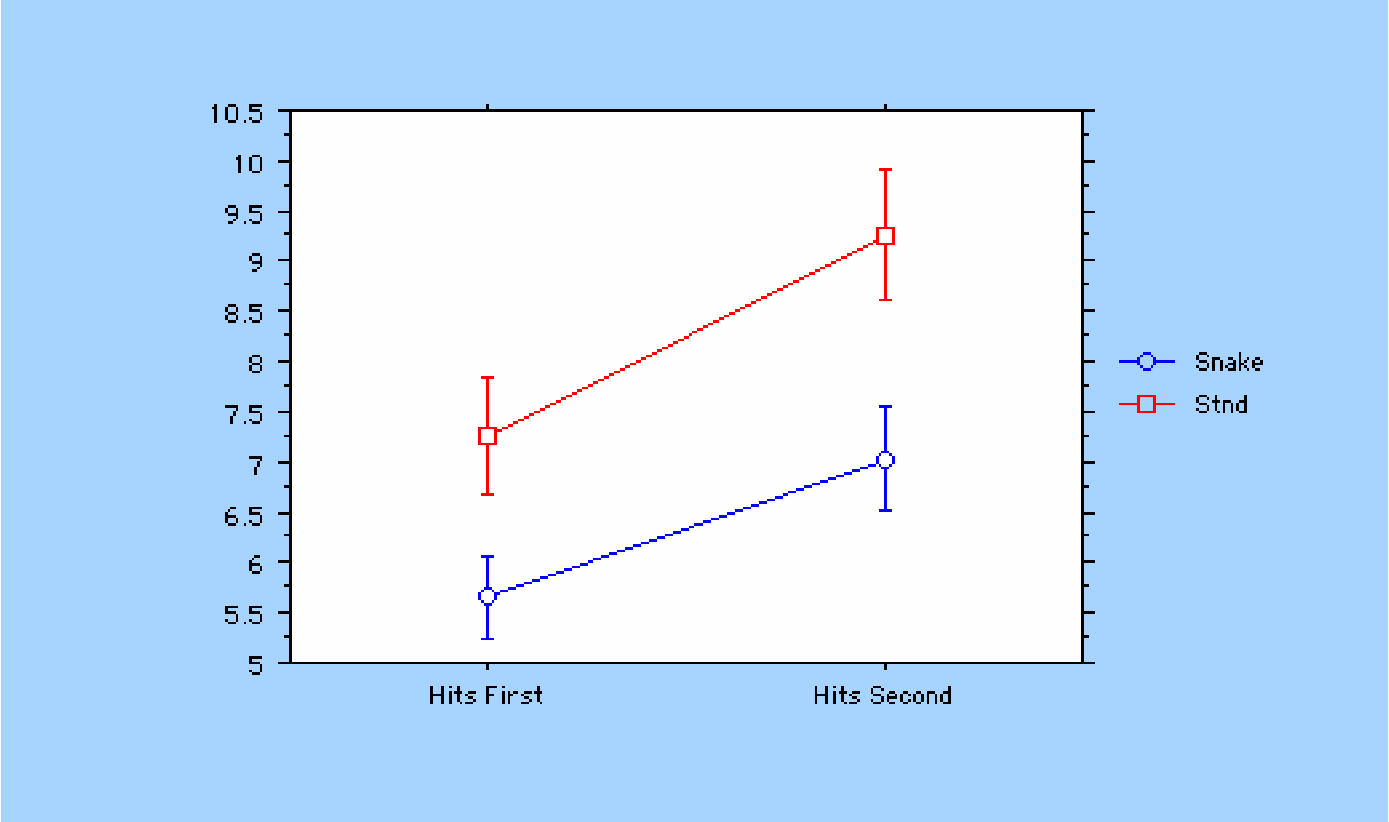
- In the “snake” condition the target production was always removed from the possible outcomes

- As a result, no success was experienced by participants even if the standard evaluation criterion was used.

The model predictions

- In the case of the “snake” condition the model predicts similar, random-level results (i.e. about five hits) for the two phases.

The results



Discussion

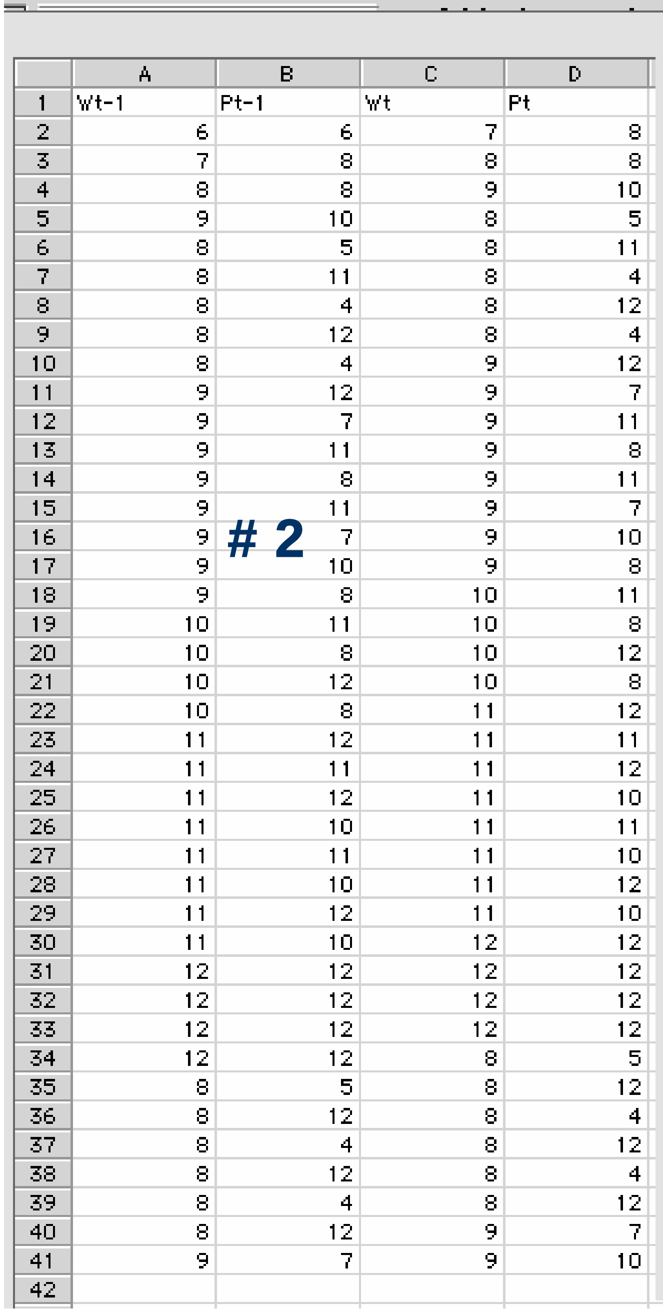
- Apparently, the participants in the “snake” condition, although performing at a lower level than those in the standard one, are able to increase their performance from the first to the second phase.

… and lead us not into temptation

- Just to immediately dispel the idea that people could approach this task by relying on the memory of previous instances, it is useful to look at the performance of individual participants.

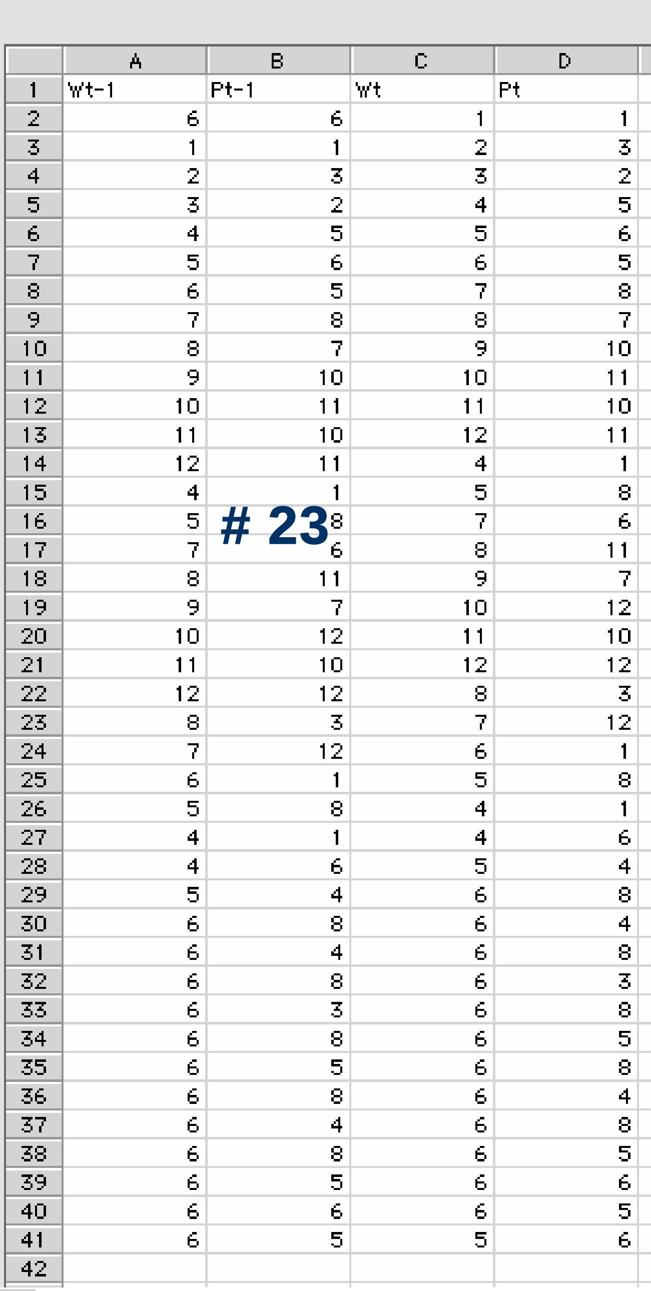
The top scorers

- These are the execution traces of the best scoring "snake" participants in the second phase.



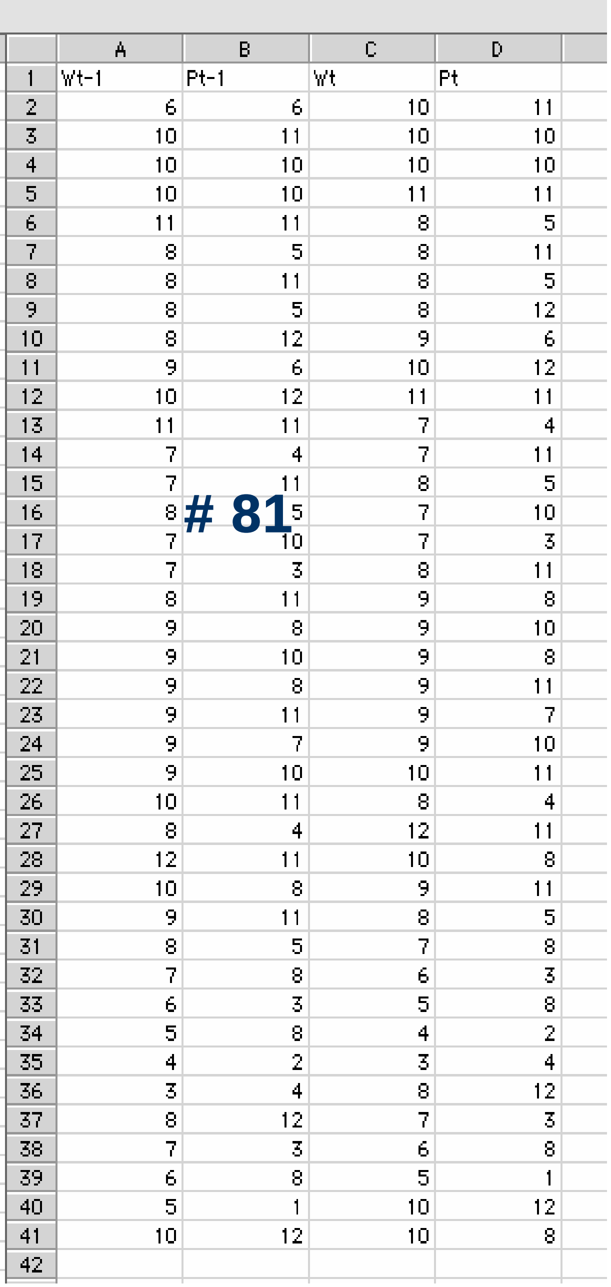
2

	A	B	C	D
1	Wt-1	Pt-1	Wt	Pt
2	6	6	7	8
3	7	8	8	8
4	8	8	9	10
5	9	10	8	5
6	8	5	8	11
7	8	11	8	4
8	8	4	8	12
9	8	12	8	4
10	8	4	9	12
11	9	12	9	7
12	9	7	9	11
13	9	11	9	8
14	9	8	9	11
15	9	11	9	7
16	9	7	9	10
17	9	10	9	8
18	9	8	10	11
19	10	11	10	8
20	10	8	10	12
21	10	12	10	8
22	10	8	11	12
23	11	12	11	11
24	11	11	11	12
25	11	12	11	10
26	11	10	11	11
27	11	11	11	10
28	11	10	11	12
29	11	12	11	10
30	11	10	12	12
31	12	12	12	12
32	12	12	12	12
33	12	12	12	12
34	12	12	8	5
35	8	5	8	12
36	8	12	8	4
37	8	4	8	12
38	8	12	8	4
39	8	4	8	12
40	8	12	9	7
41	9	7	9	10
42				



23

	A	B	C	D
1	Wt-1	Pt-1	Wt	Pt
2	6	6	1	1
3	1	1	2	3
4	2	3	3	2
5	3	2	4	5
6	4	5	5	6
7	5	6	6	5
8	6	5	7	8
9	7	8	8	7
10	8	7	9	10
11	9	10	10	11
12	10	11	11	10
13	11	10	12	11
14	12	11	4	1
15	4	1	5	8
16	5	8	7	6
17	7	6	8	11
18	8	11	9	7
19	9	7	10	12
20	10	12	11	10
21	11	10	12	12
22	12	12	8	3
23	8	3	7	12
24	7	12	6	1
25	6	1	5	8
26	5	8	4	1
27	4	1	4	6
28	4	6	5	4
29	5	4	6	8
30	6	8	6	4
31	6	4	6	8
32	6	8	6	3
33	6	3	6	8
34	6	8	6	5
35	6	5	6	8
36	6	8	6	4
37	6	4	6	8
38	6	8	6	5
39	6	5	6	6
40	6	6	6	5
41	6	5	5	6
42				



81

	A	B	C	D
1	Wt-1	Pt-1	Wt	Pt
2	6	6	10	11
3	10	11	10	10
4	10	10	10	10
5	10	10	11	11
6	11	11	8	5
7	8	5	8	11
8	8	11	8	5
9	8	5	8	12
10	8	12	9	6
11	9	6	10	12
12	10	12	11	11
13	11	11	7	4
14	7	4	7	11
15	7	11	8	5
16	8	5	7	10
17	7	10	7	3
18	7	3	8	11
19	8	11	9	8
20	9	8	9	10
21	9	10	9	8
22	9	8	9	11
23	9	11	9	7
24	9	7	9	10
25	9	10	10	11
26	10	11	8	4
27	8	4	12	11
28	12	11	10	8
29	10	8	9	11
30	9	11	8	5
31	8	5	7	8
32	7	8	6	3
33	6	3	5	8
34	5	8	4	2
35	4	2	3	4
36	3	4	8	12
37	8	12	7	3
38	7	3	6	8
39	6	8	5	1
40	5	1	10	12
41	10	12	10	8
42				

A second scenario

- In some respects, the SF paradigm seems unnatural.
- Well-behaved noise distributes normally and not step-wise, as in SF.
- What would happen if we modify the SF task to allow finely-grained (almost continuous), normally distributed noise?

A second experiment

- We carried out a second experiment in which the SF simulator generated the noise ε from a Gaussian distribution having $\mu = 0$ and $\sigma = 0.5$

- The outcome of the simulator was displayed at the unit level (e.g.: 8956 sugar tons)

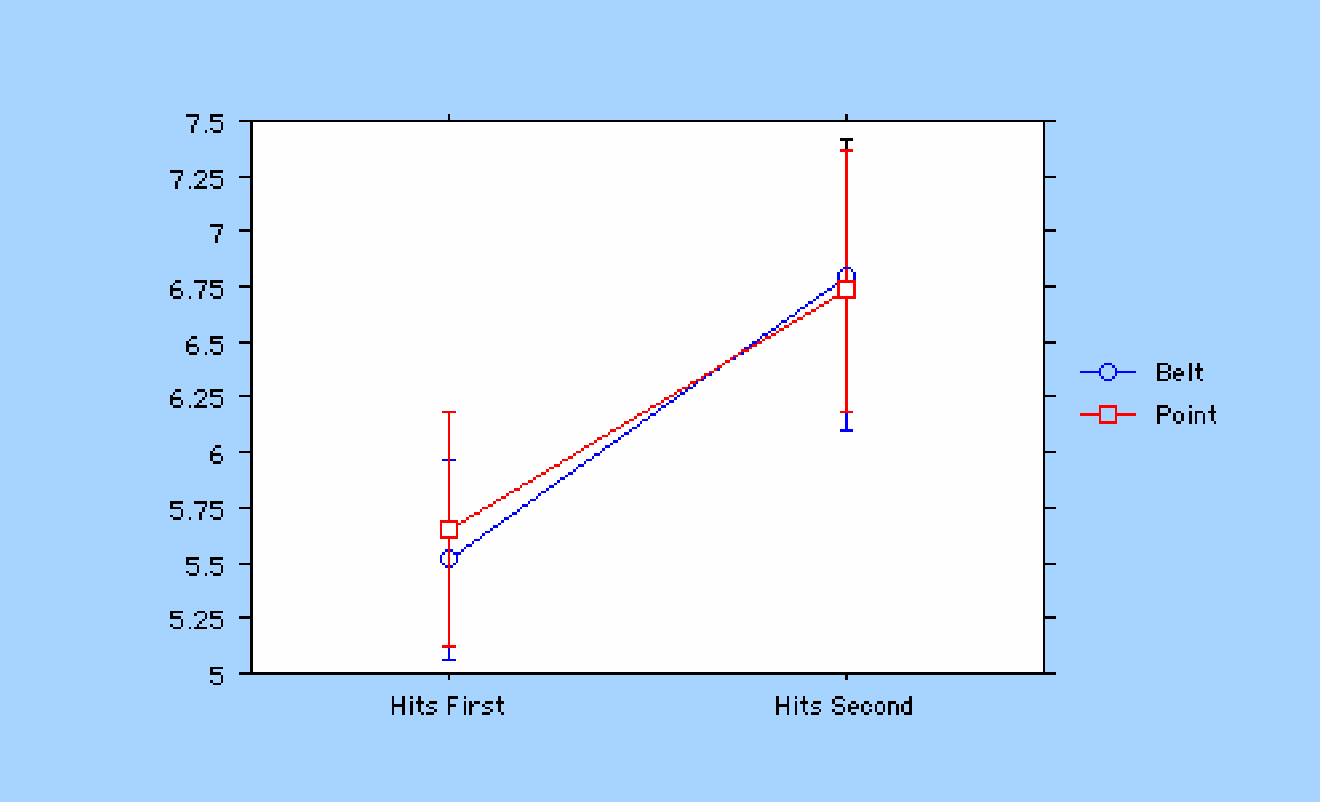
- We compared two conditions:
  (a) one in which the participants had to reach a target of exactly 9000 tons (the “point” condition)
  (b) one in which the participants had to keep the sugar production comprised between 8000 and 10000 tons (the “belt” condition).

The model predictions

- Because in the “belt” condition the productions have a higher probability of being successful, the condition will lead to better performance and substantial learning

- On the other hand, since it is almost impossible to experience success in the “point” condition, lower, random-level performance and no learning should be expected.

The results


7.5
7.25
7
6.75
6.5
6.25
6
5.75
5.5
5.25
5
Hits First
Hits Second
Belt
Point

The results

- … no difference between the conditions,
- and this leads to …

The blue part of the talk

Summing up:

- a reasonably good model capable of explaining a wide range of phenomena

- breaks downs under critical (i.e. when it is impossible or extremely unlikely to obtain a reinforcing feedback) conditions.

Some questions

- How can these findings be explained?
- How general are they?
- Do they have any implications for the ACT-R architecture?

Implications for ACT-R

These findings raise (at least) two issues concerning ACT-R:

- Is it possible to obtain subsymbolic procedural learning without experiencing overt success?
- Is success an all-or-none matter?

Generality

Some people have being exploring tasks that have a structure similar to the continuous version of SF as far as goal attainment is concerned:

- Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293-1295.

- Tomb, I., Hauser, M., Deldin, P., & Caramazza, A. (2002). Do somatic markers mediate decisions on the gambling task? *Nature Neuroscience*, 5, 1103-1104.

Possible explanations

- Several paths open to exploration (the borrowing part)
- From animal psychology/neuroscience: separation between goal and reward expectancy
  - Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to the degree of reward expectancy. *Science*, *296*, 1709-1711.
  - Richmond, B. J., Liu, Z., & Shidara M. (2003). Predicting Future Rewards. *Science, 301*, 179-181.
- From behavioral decision making: the use of a utility/ evaluation function.

A preliminary model

- Idea: the probability of an outcome to be considered as a success diminishes with the increase in its difference from the target value.

- The model has a single parameter, the standard deviation of the Gaussian function ($\mu = 0$).

Simulations Exp #1

	Stnd		Snake	
Particip.	7.26	9.26	5.66	7.03
$\sigma = 0.5$	7.36	9.19	4.57	5.03
$\sigma = 1.0$	7.81	9.80	5.09	6.14
$\sigma = 1.5$	8.01	10.23	5.45	6.78
$\sigma = 2.0$	8.18	10.38	5.49	7.02

Simulations Exp #2

	Belt		Point	
Particip.	5.51	6.80	5.65	6.74
$\sigma = 0.01$	3.47	3.89	3.58	3.86
$\sigma = 0.5$	4.38	5.35	4.37	5.43
$\sigma = 1.0$	4.75	5.92	4.76	5.88
$\sigma = 1.5$	5.06	6.19	4.94	6.04
$\sigma = 2.0$	5.09	6.18	5.02	6.06
$\sigma = 2.5$	4.98	6.08	5.15	6.21

The final slide

- Not really satisfactory results, but it’s a first step

The final slide

- Not really satisfactory results, but it’s a first step
- Still much rumination needed

The final slide

- Not really satisfactory results, but it’s a first step
- Still much rumination needed
- We have just begun to scratch the surface of what promises to be an interesting vein.