

9.591 / 24.945

Individual differences in sentence comprehension.
Working memory capacity and sentence comprehension.

September 27, 2004

Ted Gibson & Ev Fedorenko

Department of Brain & Cognitive Sciences

Department of Linguistics & Philosophy

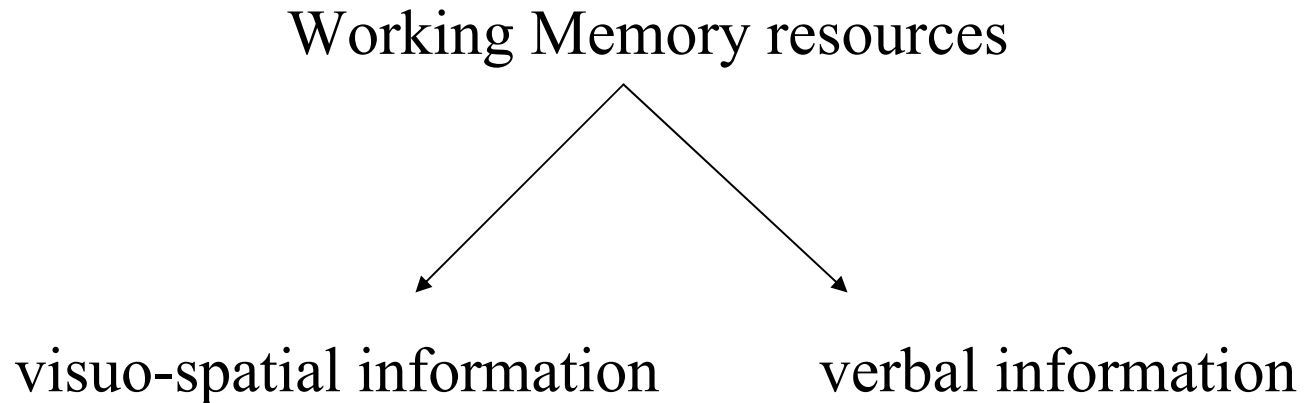
MIT

To do today:

Resources and individual differences in sentence comprehension:

1. Just & Carpenter (1992): The single resource (SR) hypothesis
2. Caplan & Waters (1999): The Separate Language Interpretation Resource (SLIR) hypothesis
3. Evidence relevant to each:
 - group differences: high and low span
 - group differences: brain damaged populations
 - dual task performance
4. An orthogonal issue: representing resources independently of language representations (MacDonald & Christianson, 2001)

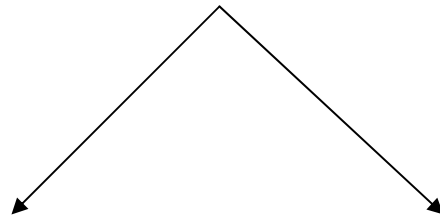
Background: Working Memory



(Baddeley & Hitch, 1974; Baddeley, 1986; Vallar & Shallice, 1990; Hanley et al., 1991; Jonides et al., 1993; Shah & Miyake, 1996)

Background: Working Memory

Verbal Working Memory resources



natural language comprehension

verbally-mediated cognitive tasks

(Caplan & Waters, 1999; cf. Just & Carpenter, 1992)

Terminological note: Working memory = computational resources

WM and sentence comprehension

Research question:

Does natural language processing rely on the same pool / overlapping pools of resources as other cognitive tasks?

Alternatively, is there a specialized pool of resources solely dedicated to language processing?

WM and sentence comprehension

Just & Carpenter (1992)

Single Resource (SR) theory: One generic pool of resources for all verbally-mediated tasks.

Caplan & Waters (1999)

Separate Language Interpretation Resource (SLIR) theory: At least two pools of resources for verbally-mediated tasks: (1) natural language processing; and (2) other non-linguistic verbally-mediated tasks.

WM and sentence comprehension: Caplan & Waters (1999)

C&W distinguish between interpretive (on-line) and post-interpretive (off-line) processes in sentence comprehension.

Definition of interpretive processes:

Automatic, first-pass language processing (“extraction of meaning from a linguistic signal”, p. 79). This includes using all sources of information.

This process breaks down on nested examples:

The man that the woman that the child hugged kissed laughed.

WM and sentence comprehension: Caplan & Waters (1999)

Definition of post-interpretive processes:

Controlled conscious processing of the propositional content of the sentence and using it to accomplish tasks, like reasoning, planning actions, storing information in long-term semantic memory, etc.

E.g.,

Pick up four tomatoes, a pound of apricots, prune juice, shallots, six apples and a bag of carrots on the way home.

The sentence above is easy to understand, but carrying out the request from memory might be hard. (But, of course, this has *nothing* to do with sentence comprehension abilities.)

WM and sentence comprehension

Traditional approaches used to investigate the question of WM resources used for natural language processing:

- **Individual differences:**

- Participants are divided into two or more groups based on their performance on a verbal WM task and then tested on linguistic structures of various complexity.

- Participants' performance on a verbal WM task is correlated with their sentence processing abilities.

- Brain damaged populations' sentence processing abilities are evaluated.

The SR theory predicts that people with different resource capacities should behave differently.

- **Dual task:**

- Participants perform two tasks: (1) on-line sentence processing; and (2) a non-linguistic verbally-mediated task (e.g. a digit span).

The SR theory predicts that we should observe a super-additive interaction when the complexity of both tasks is high.

Normal subjects: Individual differences in speed and accuracy of sentence processing

Working memory is measured using the reading span task (Daneman & Carpenter, 1980).

Reading span:

Read sentences out loud while remembering the last word of each sentence.

Ok for two sentences; Hard at 3 sentences; Very hard for 4 or more.

Average MIT undergrad reading span: 3-3.5.

Normal subjects: Individual differences in speed and accuracy of sentence processing

Variation of reading span task:

Read sentences self-paced, and answer a question about each, while keeping track of the last word of each.

Normal subjects: Individual differences in speed and accuracy of sentence processing

Possible approach:

Measure WM capacity, and look for correlations between capacity and speed/efficiency of processing.

Simplified version of this approach (used more frequently):

Divide subjects into two or three groups - High, (Medium), and Low Span - and look for group differences in comprehension speed and accuracy.

Normal subjects: Individual differences in speed and accuracy of sentence processing

Many studies don't even find RT differences between the subject groups (e.g., Caplan & Waters, 1999; Pearlmutter & MacDonald, 1995)

It is necessary to find an interaction between group-type (high, medium, low span) and sentence complexity in order to argue for the SR theory.

Processing subject- and object-extracted relative clauses (King & Just, 1991; Just & Carpenter, 1992)

Object-extraction (high syntactic complexity):

The reporter that the senator attacked admitted the error.

Subject-extraction (low syntactic complexity):

The reporter that attacked the senator admitted the error.

SR theory predicts an interaction between syntactic complexity (low, high) and group-type (low, medium, high span).

Furthermore, such an interaction should be observed in the RC region and/or the main verb region (i.e. in the regions where the syntactic complexity is manipulated), but not at the other regions.

This is a 3-way interaction: group-type (span), syntactic complexity, and sentence region.

Results: King & Just (1991); Just & Carpenter (1992)

Graphs removed for copyright reasons.

Processing subject- and object-extracted relative clauses (King & Just, 1991; Just & Carpenter, 1992)

K&J and J&C imply that there is an interaction between syntactic complexity and span: i.e., that there is a difference between H and L span subjects in the object-extracted RCs, but not in the subject-extracted RCs.

But *no required statistics are reported*. There is a main effect of span (L slower than H), and a main effect of extraction type (object-extraction slower than subject-extraction), but no interaction.

Furthermore, the RT data are from a task combining the span task (remembering the last word in sentences) and self-paced reading, so the RTs are very slow.

Processing subject- and object- extracted RCs (Caplan & Waters, 1999)

Graph removed for copyright reasons.

Caplan & Waters (1999) attempt to find the relevant interaction: None there.

The graph shows the differences in self-paced listening times for each phrase for object- minus subject-extraction sentences.

Obj-RC: The boy that the girl pushed kissed the baby.

Subj-RC: The boy pushed the girl that kissed the baby.

Ambiguity resolution: Individual differences in information use

1. Just & Carpenter (1992)

Main-verb (MV) / Reduced-relative (RR) ambiguity:

The evidence/defendant (that was) examined by the lawyer turned out to be unreliable.

Ferreira & Clifton's (1986) results: People are slow on the disambiguating by-phrase independent of the plausibility information (the animacy of the subject NP).

(Note: Trueswell, Tanenhaus & Garnsey (1994) later showed this result to be incorrect: the plausibility information of the subject NP *does affect* the ambiguity resolution difficulty.)

Ambiguity resolution: Individual differences in information use

Just & Carpenter (1992) claim that the difficulty in the by-phrase is affected by the **working memory capacity** of the reader

Results:

In the reduced form (without “that was”), low span subjects show no evidence of using plausibility information (same for both animate (“defendant”) and inanimate (“evidence”)), but high span subjects show a difference (inanimate faster than animate).

Just & Carpenter (1992) results

- Replication of F&C's results for low span subjects
- High span subjects show a different pattern

Graphs removed for copyright reasons.

Just & Carpenter (1992) results

Major problem in interpreting these results (Caplan & Waters, 1999):

Graphs removed for copyright reasons.

The high spans show the same difference between animate and inanimate in the unambiguous conditions. In order to make the first result interpretable, we would expect no differences here. This makes the results impossible to interpret.

Ambiguity resolution: Individual differences in information use

2. MacDonald, Just & Carpenter (1992)

Main-verb (MV) / Reduced-relative (RR) ambiguity:

The experienced soldiers warned about the dangers ...

MV continuation: before the midnight raid.

RR continuation: conducted the midnight raid.

For the RR continuations, there was an ambiguity effect (compared to unambiguous controls), but no group effects for L, M, H span subjects.

MacDonald, Just & Carpenter (1992) results

Unexpectedly, MJC found a span effect for the MV continuation, such that high span subjects were **slower** than low span subjects during the last word of the sentence (“raid”).

Graph removed for copyright reasons.

They argue that this effect is a result of high span subjects holding multiple interpretations in parallel, and then going more slowly.

Caplan & Waters (1999) results

But: Why does the effect only appear at the end of the sentence? Why not during the ambiguous region?

Caplan & Waters (1999) could not replicate the result.

Graph removed for copyright reasons.

Differences in self-paced RTs for ambiguous main-verb sentences minus unambiguous main-verb sentences. Region 1: “warned about the dangers”; region 2: “before the midnight”; region 3: “raid.”

Caplan & Waters (1999) results

Differences in self-paced RTs for ambiguous RR sentences minus unambiguous RR sentences. Region 1: “warned about the dangers”; region 2: “conducted the midnight”; region 3: “raid.”

Graph removed for copyright reasons.

Studies of patients with reduced working memory capacity

1. Patients with short-term memory disorders

Long term memory ok, but rehearsal, storage are impaired.

Waters et al. (1991): patient B.O. - memory span of 2 or 3 items (normal span - around 7.5); reading span of 1, but understood the sentences well.

Tested on garden-path and non-garden-path sentences. No performance differences from age-matched controls.

Waters et al. (1991): Patient B.O. (short-term memory disorder)

Three graphs removed for copyright reasons.

Studies of patients with reduced working memory capacity

2. Patients with limitations of the central executive:
Patients with **Dementia of the Alzheimer Type (DAT)** (Rochon, Caplan & Waters, 1994)

22 subjects;

2x2 design: syntactic complexity (simple, complex) x
number of propositions (1, 2);

sentence-picture matching task

Rochon, Caplan & Waters (1994): Studies of DAT patients

Results: main effects of group (DAT vs. normals) and syntactic complexity, but no interactions.

Two graphs removed for copyright reasons.

Rochon, Caplan & Waters (1994): Studies of DAT patients

Interaction in
complexity
(complex, simple)
and number of
propositions (1, 2, 3)
across both the DAT
and the control
populations.

Graph removed for copyright reasons.

Studies of patients with reduced working memory capacity

3. Patients with **Parkinson's Disease (PD)**, Waters et al. (1995)

17 subjects;

2x2 design: syntactic complexity (simple, complex) x number of propositions (1, 2);
sentence-picture matching task.

(Same design and task as for DAT subjects.)

PD patients: Waters et al. (1995)

Similar results: main effects of group (PD vs. normals) and syntactic complexity, but no interactions.

Graph removed for copyright reasons.

Interaction in group and number of propositions.

PD patients: Waters et al. (1995)

PD patients on sentence-picture matching task with and without concurrent digit load.

Graph removed for copyright reasons.

NI: No interference;

Span-1: concurrent one-less-than-span condition;

Span: concurrent span condition

Main effect of span.

Patients with reduced resources for syntactic processing: Broca's aphasics

Caplan & Waters (1996): selected 10 patients (out of 200) using a sentence-picture matching task, who met the following criteria:

- 1) showed effects of syntactic complexity;
- 2) performance above chance and below ceiling;
- 3) ability to repeat single words (necessary for the performance of a digit span task).

Patients with reduced resources for syntactic processing: Broca's aphasics

Task: sentence-picture matching

2x2 design:

(high, low syntactic complexity) x

(no-interference, concurrent load = keeping one digit in memory)

Results: main effects of syntactic complexity, interference, but no interaction

Dual task expts: The effects of external memory load on sentence processing

SR theory predicts an interaction between external load (e.g., storing a list of words) and syntactic complexity.

No such interaction is observed (Waters, Caplan & Rochon, 1995; Waters, Caplan & Hildebrandt, 1987).

Graph removed for copyright reasons.

Dual task expts: The effects of external memory load on sentence processing

King & Just (1991) manipulated:

- 1) syntactic complexity: subject / object-extracted RCs
- 2) group: H, M, L span
- 3) external load: one, two or three sentence-final words in memory, progressively through the experiment

Results:

- In recall: 1) Interaction between group and external load;
2) Interaction between syntactic complexity and external load.

But recall is, by definition, post-interpretive processing.

So these effects are not informative with regard to on-line (interpretive) sentence processing.

Dual-task: Gordon et al. 2002

2 x 2 x 2 design:

1. Load type: 3 names (e.g., Joel, Greg, Andy) or 3 descriptions (e.g., poet, cartoonist, voter)
2. Type of NP in cleft: Names vs. Descriptions
3. Sentence complexity: Subject- vs. Object-extraction in a cleft sentence:
 1. N: It was Tony that liked Joey / Joey liked before the argument began.
 2. D: It was the dancer that liked the fireman / the fireman liked before the argument began.

Gordon et al. 2002 Results

1. Interaction between sentence complexity and match / no-match in response accuracies: people were worst on the match+complex cases.
2. The on-line effect was not significant, although it was in the predicted direction.

Dual-Task Approach - Worth another try?

Previous dual-task experiments found either no on-line interaction, or only a suggestion of one (e.g. King & Just, 1991; Just & Carpenter, 1992; Caplan & Waters, 1999; Gordon et al., 2002).

But:

-secondary task usually involved storage of words/digits across the sentence processing task.

Maybe the lack of interaction is due to the fact that storage involved in language processing is *qualitatively different* from storage involved in keeping track of lists of unconnected items?

Motivations

- (1) Inconclusive results reported by King & Just (1991), Just & Carpenter (1992), and Gordon et al. (2002).
- (2) Lack of interactions in the dual-task experiments could be due to the distinctness of the storage processes involved.
- (3) No previous attempts to explore the potential interaction between the *integration processes* in sentence comprehension and secondary tasks involving similar but non-linguistic on-line integration processes.

Experiment 1: Method

- Participants - 48 subjects
- Design - 2 x 2
- Factors:
 - syntactic complexity
(subject-/object-extracted RCs);
 - arithmetic complexity
(simple additions, complex additions)

Experiment 1: Method

- Language materials - 32 sets of sentences (each with four versions as shown below)

a. Subject-extracted (easy), version 1:

The janitor who frustrated the plumber lost the key on the street.

b. Subject-extracted (easy), version 2:

The plumber who frustrated the janitor lost the key on the street.

c. Object-extracted (hard), version 1:

The janitor who the plumber frustrated lost the key on the street.

d. Object-extracted (hard), version 2:

The plumber who the janitor frustrated lost the key on the street.

Experiment 1: Method

- Math materials - randomly generated online for each participant with the following constraints:

	Initial addend	Subsequent addends
Easy math	1-10	1-3
Hard math	11-20	4-6

Experiment 1: Method

- Procedure:
 - self-paced region-by-region reading with a moving window display;
 - each sentence had four regions (2nd region was the critical region);
 - the addition task was presented above the sentence fragments simultaneously.

Sample item:
low syntactic complexity,
easy math

6

The janitor --- ----- .

+2

--- ----- who frustrated the plumber ----- --- --- --- --- -----.

+1

----- lost the key -- --- -----.

+3

----- on the street.

Type in the sum:

Q1: The janitor frustrated the plumber. T or F?

Q2: The plumber lost the key on the street. T or F?

Type in the sum:

12

Q1: The janitor frustrated the plumber. T

Q2: The plumber lost the key on the street. F

Sample item:
high syntactic complexity,
hard math

13

The employee --- --- ----- ----- ----- ----- ----- ----- ----- ----- .

+5

--- ----- who the executive praised ----- --- ----- ----- .

+6

----- finished the project -----.

+4

----- right on time.

Experiment 1: Method

Type in the sum:

Q1: The executive praised the employee. T or F?

Q2: The executive finished the project right on time. T or F?

Experiment 1: Method

Type in the sum:

28

Q1: The executive praised the employee. T

Q2: The employee finished the project right on time. T

Experiment 1: Results

- **Arithmetical accuracy**

-89.5%

-main effect of arithmetic complexity (hard math items - lower accuracies)

($F(1,47)=7.87$; $p < .01$; $F(1,31)=8.12$; $p < .01$)

Arithmetic complexity

Linguistic complexity

Subject-extraction (Easy)

Object-extraction (Hard)

Easy arithmetic

93.5 (1.7)

89.8 (1.8)

Hard arithmetic

86.7 (2.4)

87.8 (2.0)

Experiment 1: Results

- Comprehension question performance

-Question 1 - 81.5%, Question 2 - 79.4%

-main effect of syntactic complexity ($F1(147)=13.37$; $p < .001$; $F2(1,31)=6.41$; $p < .02$), main effect of arithmetic complexity in the participants analysis ($F1(1,39)=6.08$; $p < .02$; $F2(1,31)=3.52$; $p = .07$)

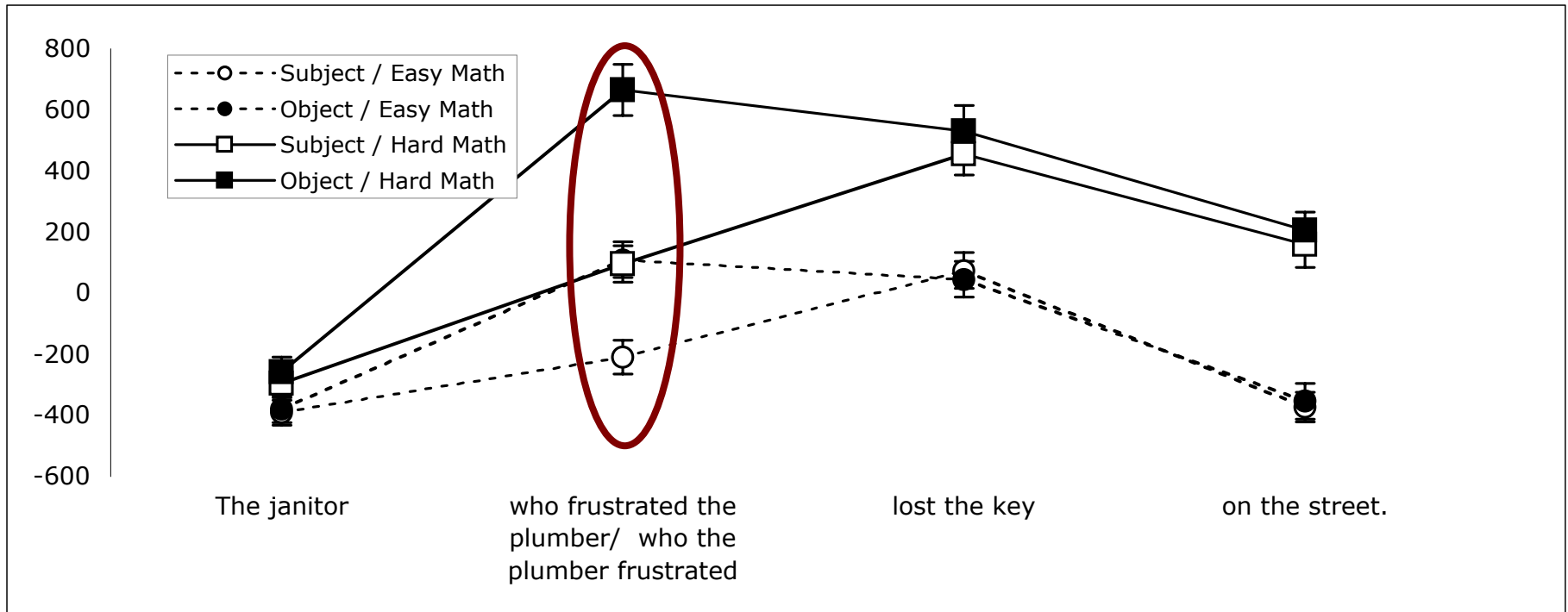
Arithmetic complexity

Linguistic complexity

	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	85.8 (2.1)	78.8 (2.7)
Hard arithmetic	80.2 (2.4)	77.0 (2.1)

Experiment 1: Results

- Reading times



Experiment 1: Results

- Reading times

-In critical region 2 we observed two main effects and a significant interaction.

-Main effect of arithmetic complexity: complex additions slower
($F(1,47)=51.53$; $p < .001$; $F(1,31)=41.67$; $p < .001$).

-Main effect of syntactic complexity: object extractions slower
($F(1,47)=36.96$; $p < .001$; $F(1,31)=43.46$; $p < .001$).

-Significant interaction

($F(1,47)=4.4$; $p < .05$; $F(1,31)=5.78$; $p < .05$).

Experiment 1: Conclusions

When both tasks are difficult (as in region 2 where both syntactic and arithmetic complexity are manipulated), there is an *additional cost* that is higher than would be predicted if the two tasks relied on independent resource pools.

On-line sentence comprehension and on-line arithmetic processing must rely on *overlapping pools of WM resources*.

Alternative Interpretation

The observed interaction could be due to *shared attentional resources* (task-switching).

- Language and arithmetic processing rely on independent WM resource pools.
- Resources are required to switch between the two tasks.
- In the difficult conditions, either (1) the switches are more difficult, or (2) more switches are required.

How to pull apart two alternative explanations?

Alternative 1: tapping non-verbal working memory

- dual-task experiment with a spatial secondary task.

Alternative 2: eliminating the dual-task component

- test subjects on math and sentence comprehension *separately*;
- divide subjects into groups based on their performance on the on-line math task;
- look for syntactic complexity by group interaction in their sentence comprehension data.

Experiment 2: Method

- Participants - 24 subjects
- Design - 2 x 2
- Factors:
 - syntactic complexity
(subject-/object-extracted RCs);
 - complexity of the spatial-rotation task
(easy/hard)

Experiment 2: Method

- Language materials - 32 sets of sentences (each with four versions as shown below) [same as in Experiment 1]

a. Subject-extracted (easy), version 1:

The janitor who frustrated the plumber lost the key on the street.

b. Subject-extracted (easy), version 2:

The plumber who frustrated the janitor lost the key on the street.

c. Object-extracted (hard), version 1:

The janitor who the plumber frustrated lost the key on the street.

d. Object-extracted (hard), version 2:

The plumber who the janitor frustrated lost the key on the street.

Experiment 2: Method

- Spatial task materials - randomly generated online for each participant with the following constraints:

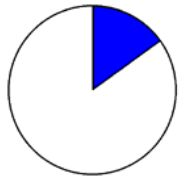
Easy condition: angles 5-90 degrees

Hard condition: angles 30-180 degrees

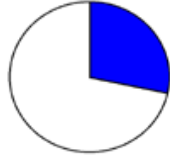
Experiment 2: Method

- Procedure:
 - self-paced region-by-region reading with a moving window display;
 - each sentence had four regions (2nd region was the critical region);
 - the spatial-rotation task was presented above the sentence fragments simultaneously.

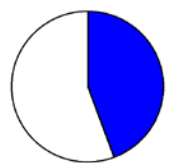
Sample item:
low syntactic complexity,
easy spatial task



The janitor -----.



--- ----- who frustrated the plumber ----- .

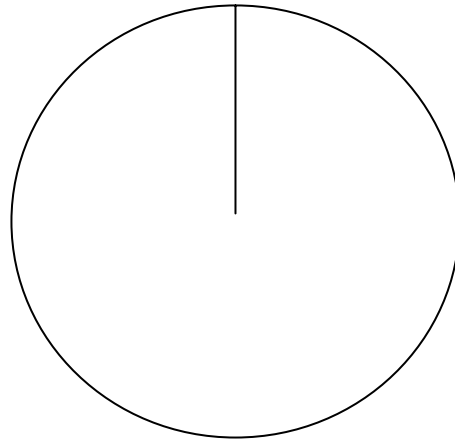


----- lost the key -----.



----- on the street.

Move the mouse to the final
summed location direction:



Q1: The janitor frustrated the plumber. T or F?

Q2: The plumber lost the key on the street. T or F?

Experiment 2: Results

- Performance on the spatial task

-30.3 degrees off from the correct answer

-main effect of complexity of the spatial task (hard spatial task items - lower accuracies)

($F(1,23)=18.36$; $MSe=2676$; $p < .001$; $F(1,31)=22.28$; $MSe=3568$; $p < .001$)

Spatial task complexity

Linguistic complexity

Subject-extraction (Easy)

Object-extraction (Hard)

Easy rotations

24.5 (2.7)

25.6 (2.2)

Hard rotations

36.4 (3.3)

34.8 (3.1)

Experiment 2: Results

- Comprehension question performance
 - Question 1 - 82.2%, Question 2 - 83.7%
 - no significant effects/interactions ($F_s < 1$)
-

Spatial task complexity

Linguistic complexity

Subject-extraction (Easy)

Object-extraction (Hard)

Easy rotations

83.3 (3.3)

83.9 (3.0)

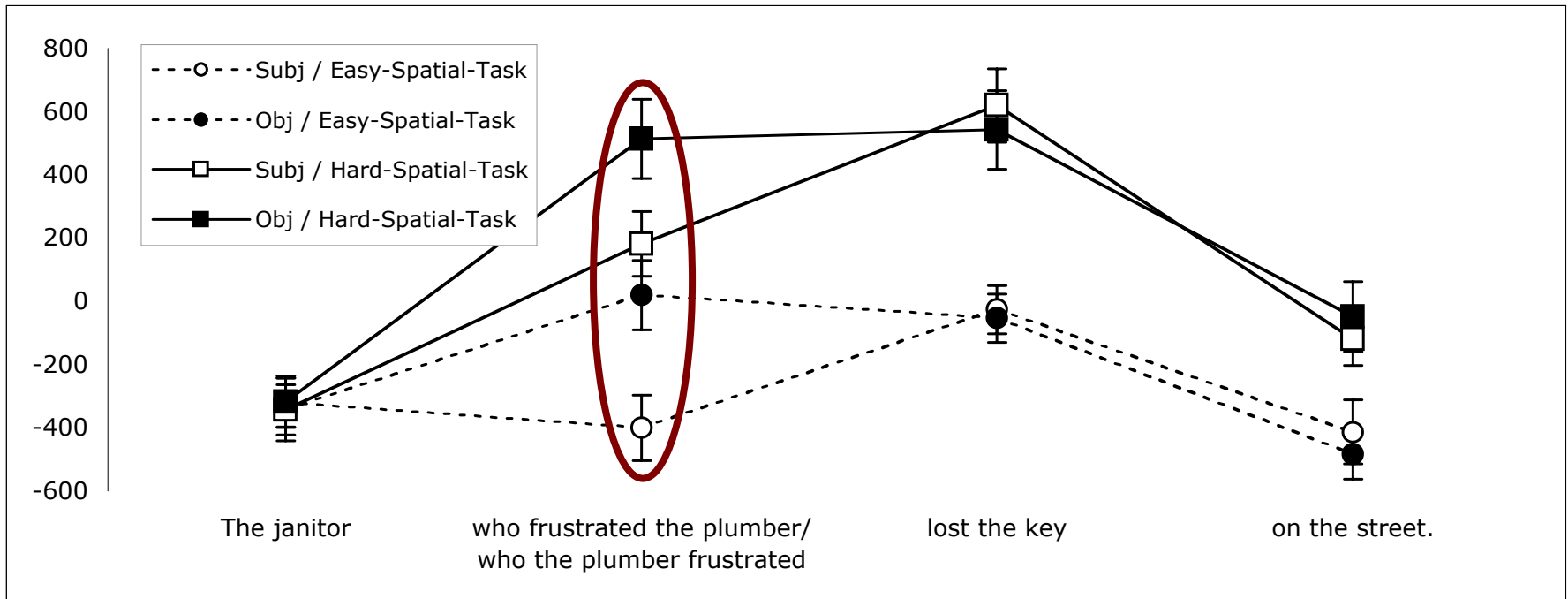
Hard rotations

81.3 (3.6)

83.3 (2.8)

Experiment 2: Results

- Reading times



Experiment 2: Results

- Reading times

- In critical region 2 we observed two main effects but NO interaction.

- Main effect of spatial task complexity: hard spatial task conditions slower
($F1(1,23)=22.02$; $p < .001$; $F2(1,31)=44.26$; $p < .001$).

- Main effect of syntactic complexity: object extractions slower
($F1(1,23)=12.84$; $p < .002$; $F2(1,31)=19.11$; $p < .001$)

Experiment 2: Conclusions

The lack of an interaction argues against the attentional account of the interaction observed in Expt 1.

On-line sentence comprehension and on-line arithmetic processing must rely on *overlapping pools of WM resources*.

Open Questions

- What is the *exact nature of the overlap* in WM resources for sentence comprehension and arithmetic processing?
- What other cognitive systems might rely on the WM resources used in on-line sentence comprehension and on-line arithmetic additions? Processing musical syntax?
- What are the implications of these results for the functional architecture of the WM system, in general?

Experiments 1 and 2: a possible re-interpretation of the results

Maybe the difference between the patterns of the results in Expts 1 vs. 2 is not due to the *verbal/spatial* difference in the nature of the secondary task, but rather due to *discrete/approximate* difference between the arithmetic/pie tasks.

Thus, the interaction in Expt 1 could be resulting from the fact that both tasks tap a WM resource pool dedicated to processing *sequential discrete* information units.

Dual task example 1

The

reporter

who

the

senator

attacked

admitted

the

error.

MacDonald & Christiansen (2001)

Another approach to working memory and language processing

Central claim: There are no “resources” for processing separate from the “resources” that are used to represent the linguistic knowledge.

This is more of a “position” paper rather than an empirical paper. This paper argues against the hypothesis that working memory resources are independent of representations.

E.g., M&C are contra Just & Carpenter, whose model includes units for representations that are independent of the units that are used in processing.

MacDonald & Christiansen (2001)

It is not clear how this model is relevant to Caplan & Waters' claims.

(a) no explicit model in C&W

(b) M&C's model may be compatible with C&W's hypotheses

MacDonald & Christiansen (2001)

Evidence:

(1) Adults perform better and faster on more complex constructions (object-extracted RCs) than on simpler constructions (subject-extracted RCs) when they are trained on the complex construction.

(2) Their connectionist simulation performed similarly.

MacDonald & Christiansen (2001)

Is this evidence for an SR theory? An SLIR theory?

It's not evidence for either. It's evidence for some role of experience in language processing.

But it's not evidence against the existence of built-in constraints as well: E.g., people get faster overall, but their behavior is still predicted by locality considerations.

Indeed, M&C's model has lots of built-in constraints: memory based.