



U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND – ARMY RESEARCH LABORATORY

Automated information extraction to facilitate comprehension
across text difficulty levels

Erin Zaroukian

Cognitive Scientist

Content Understanding Branch, Computational and Information Sciences Directorate

DISTRIBUTION UNLIMITED



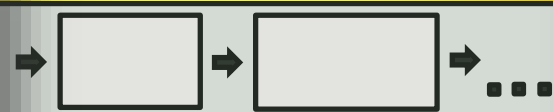
INTRODUCTION



- Information Extraction (IE) pipelines can aid decision making by structuring data and pulling decision-relevant information from large document sets
- Much research focuses on precision/recall of the pipeline
- Little research on how useful the pipeline output is to a user
 - Does text markup from an IE pipeline improve human comprehension of text documents?

All the military bases in Perchland are heavily protected.
 There is no new information about Raven group operations in Bassland.
 Perchland is land locked.
 Locals in Sharkland are being recruited.
 The Turtle lost his right eye in an accident.
 The Bronco group does not attack on its Sabbath.
 Members of the Charger, Titan, Steeler, Raider, and Bronco groups have experience with chemical weapons.
 The shopping malls in the coalition area are not well defended.
 Charger and Titan group members have entered Perchland and Salmonland.
 The Panther, Charger, Titan, and Raven groups prefer to attack in daylight.

IE Pipeline



All the [org:military] [fac:bases] in [cpe:Perchland] are heavily protected.
 There is no new information about Raven group operations in Bassland.
 Perchland is land locked.
 [per:Locals] in [cpe:Sharkland] are being recruited.
 The [per:Turtle] <lost> [per:his] right eye in an accident.
 The Bronco [org:group] does not attack on its Sabbath.
 [per:Members] of the [org:Charger], [org:Titan], [org:Steeler], [org:Raider], and [org:Bronco] [org:groups] have experience with chemical weapons.
 The shopping [fac:malls] in the coalition area are not well defended.
 Charger and Titan [org:group] [per:members] have <entered> Perchland and Salmonland.
 The [per:Panther], [per:Charger], [org:Titan] and Raven [org:groups] prefer to attack in daylight.

Event ID: EV32
 Trigger: entered
 Event Type: Movement
 Event Subtype: Transport
 Genericity: Specific
 Modality: Asserted
 Polarity: Positive
 Tense: Past
 Arguments:
 Artifact members Destination Salmonland

Unstructured
Data

Output

Analysts

Decision
Maker



EXPERIMENT 1 - RESEARCH QUESTION



Does text markup from an IE pipeline improve human comprehension of text documents?



EXPERIMENT 1 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

- ELICIT: Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust [Ruddy 2007]
- Scenario
 - 68 sentences
 - Together provide *who/what/where/when* of an anticipated adversary attack

Example mini-scenario

The Tetrahedron group is not involved.

There will be a suicide bomb attack at a new hotel.

The Cube and Tetrahedron groups have the capacity to operate in Gibbonland and Lemurland.

The Cube group only attacks on New Year's Day.

A hotel was recently built in Gibbonland.

The attack will be at 11:00 pm.

Time remaining: 19:31

Terrorist plot:

Who?

What?

Where?

When? at

[Click here to submit answers for this scenario](#)



EXPERIMENT 1 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

- Text/problem difficulty – Alston (2010), Morton & Adams (2010)
 - Amount of “noise”
 - Number of sentences required to solve
 - Number of possible solutions

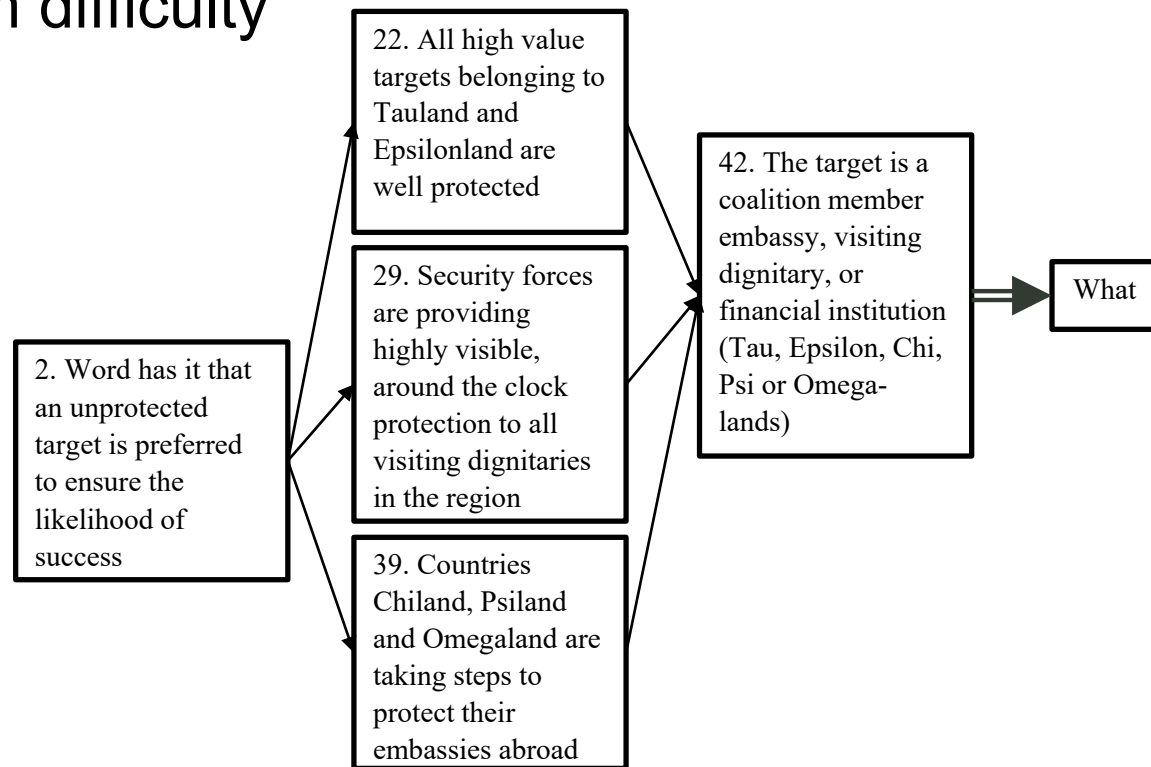


EXPERIMENT 1 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

– Text/problem difficulty



Logic chains for Scenario 1 "What", as given in Morton & Adams.



EXPERIMENT 1 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

– Text/problem difficulty

	Factoid Set 1	Factoid Set 2	Factoid Set 3	Factoid Set 4
Mixed Logic Streams	7	8	9	4
Factoids per sub-solution – <i>from Alston</i> Who, What, Where, When (Sum)	5, 5, 5, 9 (24)	5, 11, 8, 10 (34)	10, 8, 14, 4 (36)	5, 7, 6, 4 (22)
Factoids per sub-solution – <i>from Morton & Adams</i> <i>solving matrices</i>	5, 4, 7, 9 (25)	5, 5, 4, 9 (23)	8, 4, 7, 8 (27)	6, 4, 12, 8 (30)
Factoids per sub-solution – <i>from Morton & Adams</i> <i>solution trees</i>	5, 5, 8, 9 (27)	5, 6, 5, 9 (25)	11, 5, 8, 9 (32)	6, 6, 11, 9 (32)
Number of relationships – <i>from Alston</i>	25	25	27	17
Number of factoids – <i>from Alston</i>	15	15	16	12
Number of factoids – <i>from Morton & Adams</i> <i>solution trees</i>	16	17	16	17



EXPERIMENT 1 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

- Text/problem difficulty
 - Alston:
 - $4 \ll 2 < 3$
 - $4 \ll 1$
 - Morton & Adams:
 - $1 \approx 2$
 - 3 has greater dependence on interim conclusions regarding the answers to *What* and *Where*.
 - 4 is substantially different from the other three scenarios in structure
- **$4/8 < 1, 2 < 3/7$**



EXPERIMENT 1 - METHOD



Does text markup from an IE pipeline improve human comprehension of text documents?

- Markup generated by an RPI pipeline [Li, Ji, and Huang 2013; Li and Ji 2014]
 - Events, labeled entities, mouse-over

Plain scenario excerpt

All the military bases in Perchland are heavily protected.
 There is no new information about Raven group operations in Bassland. Perchland is land locked.
 Locals in Sharkland are being recruited.
 The Turtle lost his right eye in an accident.
 The Bronco group does not attack on its Sabbath.
 Members of the Charger, Titan, Steeler, Raider, and Bronco groups have experience with chemical weapons.
 The shopping malls in the coalition area are not well defended.
 Charger and Titan group members have entered Perchland and Salmonland.
 The Panther, Charger, Titan, and Raven groups prefer to attack in daylight.

Markup scenario excerpt

All the [ORG military] [FAC bases] in [CPE Perchland] are heavily protected.
 There is no new information about Raven group operations in Bassland. Perchland is land locked.
 [PER Locals] in [CPE Sharkland] are being recruited.
 The [PER Turtle] <lost> [PER his] right eye in an accident.
 The Bronco [ORG group] does not attack on its Sabbath.
 [PER Members] of the [ORG Charger], [ORG Titan], [ORG Steeler], [ORG Raider], and [ORG Bronco] [ORG groups] have experience with chemical weapons.
 The shopping [FAC malls] in the [CPE coalition area] are not well defended.
 Charger and Titan [ORG group] [PER members] have <entered> Perchland and [LOC Salmonland].
 The [PER Panther], [PER Charger], [ORG Titan] and Raven [ORG groups] prefer to <attack> in daylight.

Event ID: EV32
 Trigger: entered
 Event Type: Movement
 Event Subtype: Transport
 Genericity: Specific
 Modality: Asserted
 Polarity: Positive
 Tense: Past
 Arguments:
 Artifact members Destination Salmonland



EXPERIMENT 1 - METHOD



Does text markup from an IE pipeline improve human comprehension of text documents?

– Measured objectively as the accuracy and speed with which participants answer questions about the text

– Measured subjectively through ratings of workload and preference

[Type Mention]: Entity
 [Type Mention]: Entity with relation
 <Trigger>: Event
 [PER]: Person [ORG]: Organization [CPE]: Geo-political Entity [FAC]: Facility
 [VEH]: Vehicle [WEA]: Weapon [LOC]: Location [OO]: Others
 [TIMEX]: Timex [VALUE]: Value

The [per Eagle] is involved.
 The Eagle does not work in [cpe Spiderland].
 The northern cruise [fac terminal] is ocean-based.
 The <attack> will be at the end of the second shift.
 The [per Circle] [per group] is recruiting [per locals] – intentions unknown.
 The largest [org museum] in [cpe Spiderland] has a flat roof.
 The [per attackers] are focusing on a high visibility target.
 The [org Oval] and [per Hexagon] [org groups] want to <attack> the interests of [per Hornetland].
 The [per Heron] was <born> in [cpe Hornetland].

Time remaining: 19:54

Terrorist plot:
 Who?
 What?
 Where?
 When? at

[Click here to submit your answer and continue](#)

Use an intermediate value if you are unsure or have mixed feelings.

	Definitely with markup	Equal	Definitely without markup
	-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10		
Which version of the task felt more mentally demanding?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which version of the task felt more physically demanding?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which version of the task felt more hurried or rushed?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
On which version of the task do you think you performed better?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
On which version of the task did you feel you had to work harder?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which version of the task lead you to feel more insecure, discouraged, irritated, stressed, or annoyed?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, which version of the task do you prefer?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Click here to continue](#)



EXPERIMENT 1 – PARTICIPANTS AND PROCEDURE



– Participants

- 100 Turkers
- \$2

– Procedure

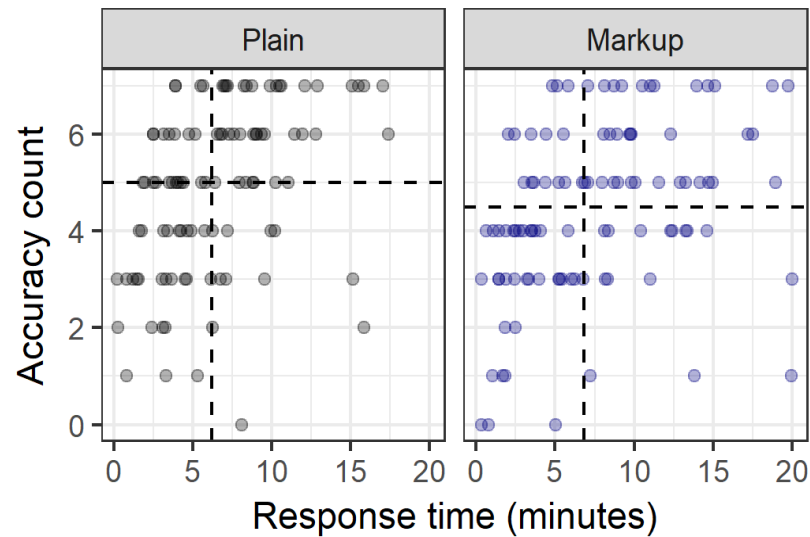
- Demographic questionnaire
- Instructions
- Plain: Practice scenario, test scenario, answers
- Markup: Practice scenario, test scenario, answers
- Workload and preference questionnaire

Condition order
randomized





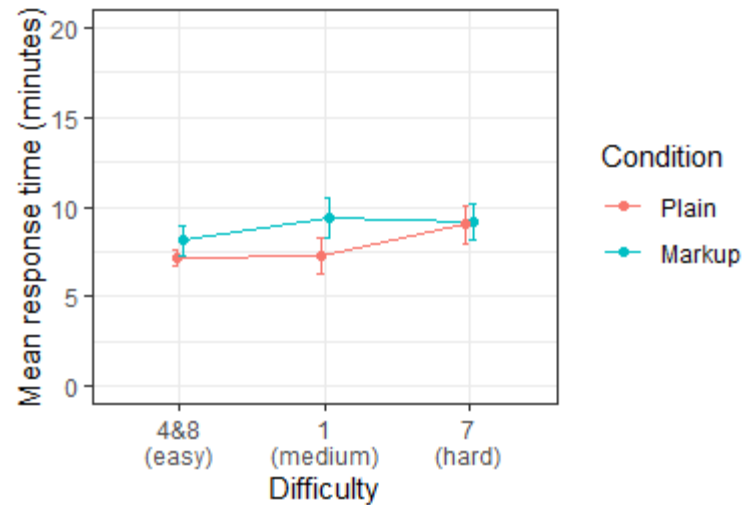
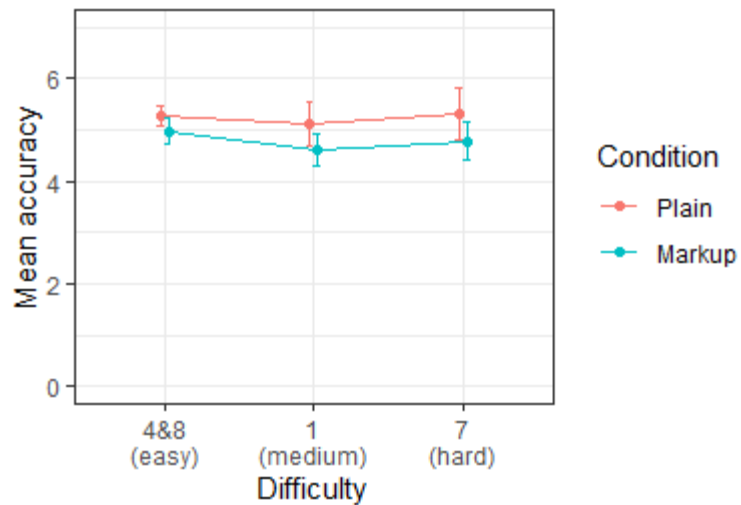
EXPERIMENT 1 - RESULTS



- **Accuracy**
 - Plain $>_*$ Markup
- **Speed**
 - Plain $<_*$ Markup
- **Workload**
 - Plain $<_*$ Markup
- **Preference**
 - Plain $>_*$ Markup



EXPERIMENT 1 - RESULTS



- **Scenario difficulty**
 - Did not affect accuracy, response time



EXPERIMENT 2 – RESEARCH QUESTION



Does text markup from an IE pipeline improve human comprehension of text documents?

Experiment 1 - comprehension *worsened* with markup, no effect of difficulty

Experiment 2 – will comprehension improve with “ideal” markup?



EXPERIMENT 2 - METHOD



Does **text** markup from an IE pipeline improve human comprehension of **text documents**?

- ELICIT: Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust [Ruddy 2007]
- Scenario
 - 68 sentences
 - Together provide *who/what/where/when* of an anticipated adversary attack

Example mini-scenario

The Tetrahedron group is not involved.

There will be a suicide bomb attack at a new hotel.

The Cube and Tetrahedron groups have the capacity to operate in Gibbonland and Lemurland.

The Cube group only attacks on New Year's Day.

A hotel was recently built in Gibbonland.

The attack will be at 11:00 pm.

Time remaining: 19:31

Terrorist plot:

Who?

What?

Where?

When? at

[→ Click here to submit answers for this scenario](#)



EXPERIMENT 2 - METHOD



Does text markup from an IE pipeline improve human comprehension of text documents?

- Hand-generated “ideal” markup
 - Potential *Who, What, Where, When* highlighted

Plain scenario excerpt

The Lion is involved.

Word has it that an unprotected target is preferred to ensure the likelihood of success.

The Lion doesn't operate in Chiland.

The Lion attacks in daylight.

All of the members of the Azure group are now in custody.

Security forces are providing highly visible, around the clock protection to all visiting dignitaries in the region.

The Lion is planning something in April on the anniversary of his father's death.

The Brown group is recruiting locals - intentions unknown.

Markup scenario excerpt

The Lion is involved.

Word has it that an unprotected target is preferred to ensure the likelihood of success.

The Lion doesn't operate in Chiland.

The Lion attacks in daylight.

All of the members of the Azure group are now in custody.

Security forces are providing highly visible, around the clock protection to all visiting dignitaries in the region.

The Lion is planning something in April on the anniversary of his father's death.

The Brown group is recruiting locals - intentions unknown.



EXPERIMENT 2 - METHOD



Does text markup from an IE pipeline improve human comprehension of text documents?

- Measured objectively as the accuracy and speed with which participants answer questions about the text

- Measured subjectively through ratings of workload and preference

The Lion is involved.

Word has it that an unprotected target is preferred to ensure the likelihood of success.

The Lion doesn't operate in Chiland.

The Lion attacks in daylight.

All of the members of the Azure group are now in custody.

Security forces are providing highly visible, around the clock protection to all visiting dignitaries in the region.

The Lion is planning something in April on the anniversary of his father's death.

The Brown group is recruiting locals - intentions unknown.

Time remaining: 19:51 may be involved.

Terrorist plot:

Who? -- Select group --

What? -- Select target --

Where? -- Select country --

When? -- Select month --

-- Select day --

at -- Select time --

-- Select AM/PM --

[Click here to submit answers for this scenario](#)

	1	2	3	4	5	6	7	Agree
The system is deceptive.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system behaves in an underhanded manner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am suspicious of the system's intent, action, or outputs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am wary of the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system's actions will have a harmful or injurious outcome.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am confident in the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system provides security.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system has integrity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system is dependable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The system is reliable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can trust the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am familiar with the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



EXPERIMENT 2 – PARTICIPANTS AND PROCEDURE



– Participants

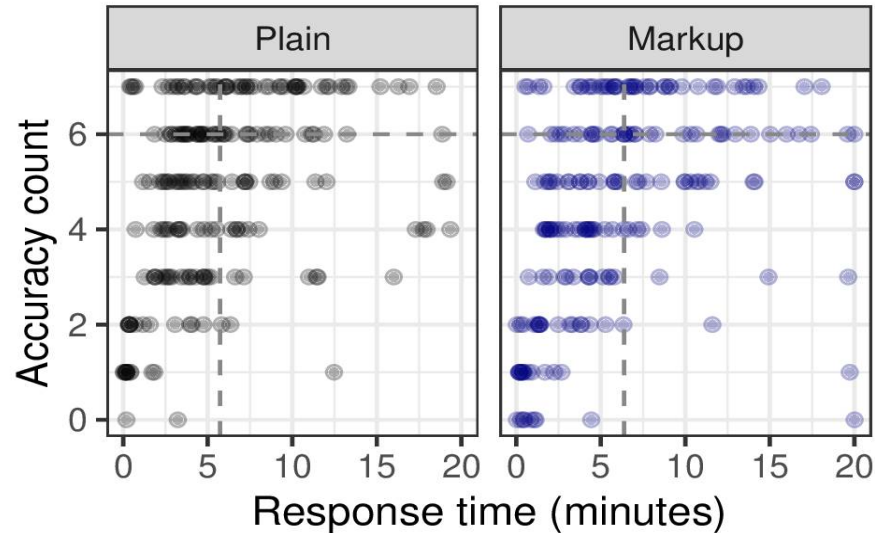
- 200 Turkers
- \$2

– Procedure

- Demographic questionnaire
- Instructions
- Plain practice scenario, Markup practice scenario
- Trust in automation survey
- Plain or Markup: Test scenario, answers, text scenario, answers
- Trust in automation survey
- Workload and preference questionnaire



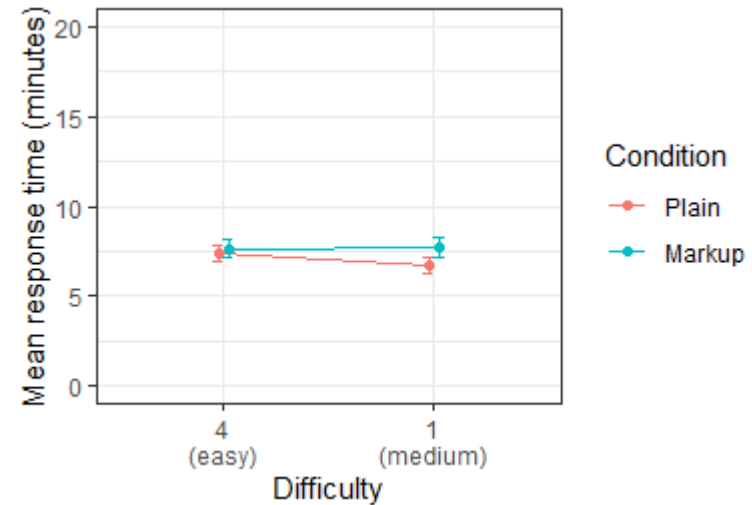
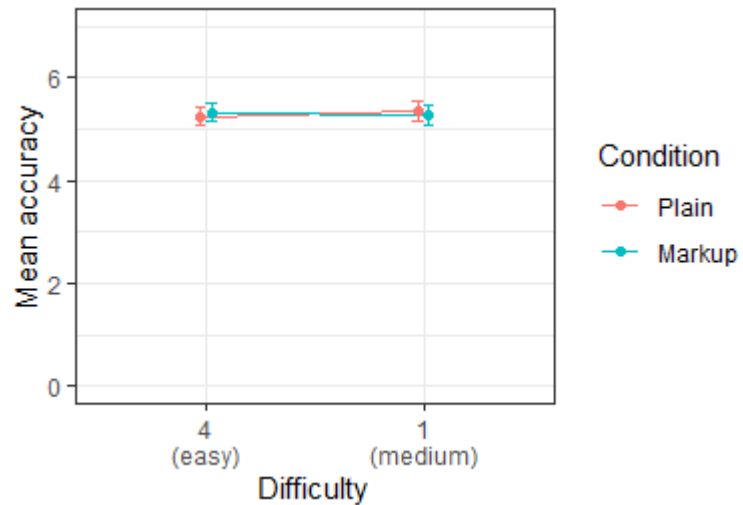
EXPERIMENT 2 – RESULTS



- **Accuracy**
 - Plain \approx Markup
- **Speed**
 - Plain \approx Markup
- **Workload**
 - Plain \approx Markup, except Overall Performance favored Markup
- **Preference**
 - Plain $<_*$ Markup



DIFFICULTY



- **Scenario difficulty**
 - Did not affect accuracy, response time



CONCLUSION



- **Experimental framework**
 - Twice failed to show accuracy/speed advantage for markup
 - Showed lower workload, preference for “ideal” markup
 - Can be used to explore further manipulations, lead to a better understanding of how various features of tasks and text presentation affect various aspects of performance.
 - E.g., task difficulty – multidimensional!
- **Specific use cases to be tested as IE pipelines are developed or as the end user’s task changes → IE development loop that includes user testing and user-directed IE development guidelines, promoting systems that succeed not only on intrinsic measures, but on extrinsic measures as well.**



THANKS!



- **Sue Kase, Michelle Vanni, Justine Caylor, Stephen Tratz, Claire Bonial, Jeffrey Micher, Clare Voss, Lucia Donatelli, Jon Bakdash of ARL**



REFERENCES



- Cunningham, H.: "Information Extraction, automatic." In: Encyclopedia of Language and Linguistics, 2nd Ed., pp. 665-677. Elsevier, New York (2005).
- Ruddy, M.: ELICIT – The Experimental Laboratory for the Investigation of Collaboration, Information Sharing, and Trust. In: Proceedings of the 12th annual International Command and Control Research and Technology Symposium (2007).
- Alberts, D.: The Agility Advantage: a survival guide for complex enterprises and endeavors. The Command and Control Research Program Publication, Department of Defense, Washington, DC (2011).
- Alston, A.: Assessing the Difficulty and Complexity of ELICIT Factoid Sets. In Proceedings of the 15th annual International Command and Control Research and Technology Symposium (2010).
- Morton, A., and B. D. Adams.: Development of a Team Scenario Content Generation Framework. No. DRDC-TORONTO-CR-2010-105. Humansystems Inc, Guelph, Ontario (2010).
- Zaroukian, E.: Information extraction for optimized human understanding and decision making. In Proceedings of the 23rd International Command and Control Research and Technology Symposium (2018).
- Li, Q., Ji, H.: Incremental joint extraction of entity mentions and relations. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 402–412. ACL, New York (2014).
- Li, Q., Ji, H., Huang, L.: Joint event extraction via structured prediction with global features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 73–82. ACL, New York (2013).
- NASA: NASA Task Load Index (TLX), v. 1.0 Manual. (1986).
- Bürkner, P. and M. Vuorre. "Ordinal regression models in psychology: a tutorial." *Advances in Methods and Practices in Psychological Science* 2.1 (2019): 77-101.
- Marusich, L. R., J. Z. Bakdash, E. Onal, M.S. Yu, J. Schaffer, J. O'Donovan, T Höllerer, N. Buchler, C. Gonzalez.: "Effects of information availability on command-and-control decision making: performance, trust, and situation awareness." *Human factors* 58.2 (2016): 301-321.
- Zaroukian, E., J. Caylor, M. Vanni, and S. Kase: Evaluating Improvement in Situation Awareness and Decision-Making through Automation. In Proceedings of the 24rd International Command and Control Research and Technology Symposium (2019).
- Zaroukian, E., J. Caylor, M. Vanni, and S. Kase: Human Interacting with the output of information extraction systems. Proceedings of the 10th International Conference on Applied Human Factors and Ergonomics. (2019).