



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

Human Interaction with the Output of Information Extraction Systems

Erin Zaroukian

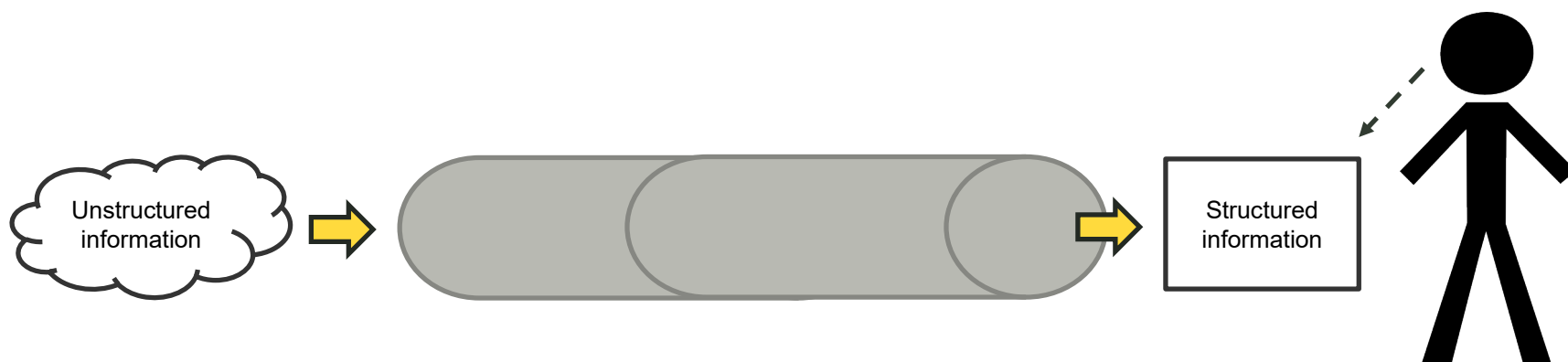
Cognitive Scientist

Computational and Information Sciences Directorate

Approved for public release, distribution
unlimited



INFORMATION EXTRACTION





Experiment 1

Markup from an IE pipeline

All the [ORG military] [FAC bases] in [GPE Perchland] are heavily protected.
 There is no new information about Raven [ORG groups] in [GPE Perchland].
 Perchland is land locked.
 [PER Locals] in [GPE Sharkland] are being hunted.
 The [PER Turtle] <lost> [PER his] right leg.
 The Bronco [ORG group] does not <attack> [LOC Salmonland].
 [PER Members] of the [ORG Charger], [ORG Bronco] [ORG groups] have experienced.
 The shopping [FAC malls] in the [GPE Perchland] are heavily 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 - PROCEDURE



Demographics

Do you consider English your first or primary language?

- ☐ Yes
- ☐ No

Please select your age range:

- ☐ 18-29
- ☐ 30-49
- ☐ 50-64
- ☐ 65+

Please select your gender:

- ☐ Female
- ☐ Male
- ☐ Other/prefer not to say

Please select your highest level of educational attainment:

- ☐ High school graduate or less
- ☐ Some college
- ☐ College degree or more



EXPERIMENT 1 - PROCEDURE



Scenario text

- With markup -

All the [ORG **military**] [FAC **bases**] in [GPE **Perchland**] are heavily protected.
 There is no new information about Raven group operations in Bassland.
 Perchland is land locked.
 [PER **Locals**] in [GPE **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 **Steelers**], [ORG **Raiders**], and [ORG **Bronco**] [ORG **groups**] have experience with chemical weapons.
 The shopping [FAC **malls**] in the [GPE **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**

⋮

Time remaining: 19:42

Terrorist plot:

Who?
 What?
 Where?
 When? at

- Without markup -

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.

⋮

Time remaining: 19:42

Terrorist plot:

Who?
 What?
 Where?
 When? at



EXPERIMENT 1 - PROCEDURE

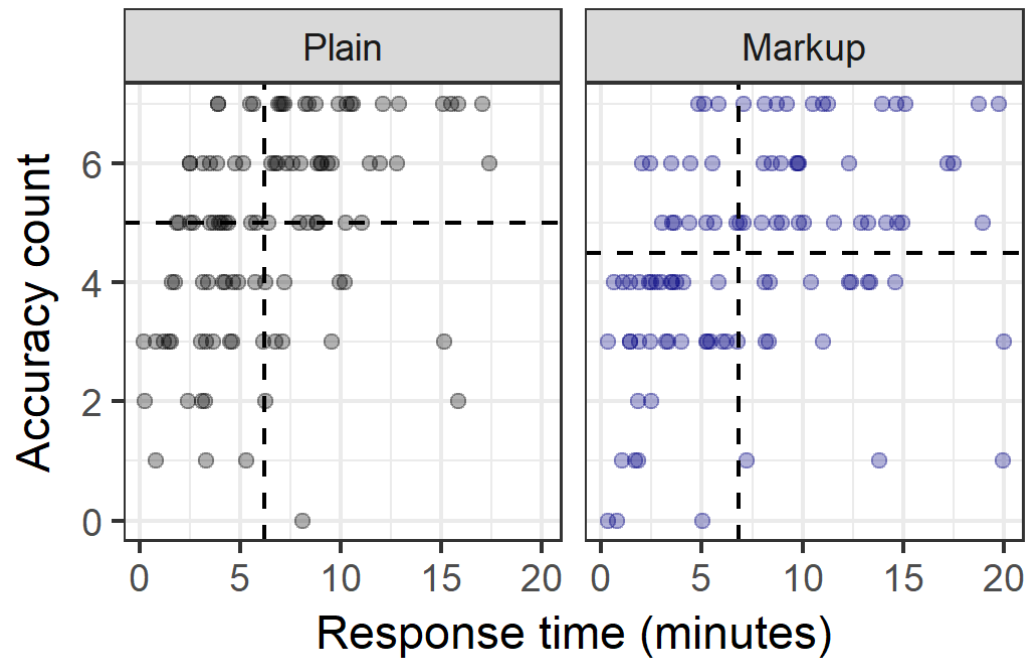


Workload (NASA-TLX) and preference

	Definitely without markup										Equal										Definitely with 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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<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 1 - RESULTS



- **Plain more accurate, faster**
 - Note – sizable minority of participants prefer markup, perform better with markup
- **Lower workload ratings for Plain**
- **Prefer Plain**



Experiment 2

“Ideal” markup

The Gray and Teal groups do not employ suicide bombers.

The Turquoise group focuses on destroying energy infrastructure.

No attacks are being planned on religious organizations in Sigmaland.

The target is in a coalition country (Mulan, Xiland, Omicronland, Piland, or Sigmaland).

An attack is being planned for the first month of the year.

The Rose group may be involved.

There is a lot of activity involving the Rose group.

Embassies in Piland were recently attacked and evidence of more attacks has been found.

There are reports that spent nuclear fuel is missing in Muland.

The largest bank in Piland has 4 machine gun emplacements on its roof.

No traces of members from the Blue group have been found in Sigmaland.

Mulan is in the process of deploying more troops to protect its military bases.



EXPERIMENT 2 - PROCEDURE



Demographics

Do you consider English your first or primary language?

- ☐ Yes
- ☐ No

Please select your age range:

- ☐ 18-29
- ☐ 30-49
- ☐ 50-64
- ☐ 65+

Please select your gender:

- ☐ Female
- ☐ Male
- ☐ Other/prefer not to say

Please select your highest level of educational attainment:

- ☐ High school graduate or less
- ☐ Some college
- ☐ College degree or more

Please select your occupation or field of study:

- ☐ Science and Technology
- ☐ Arts, Entertainment, and Media
- ☐ Education
- ☐ Legal
- ☐ Sales
- ☐ Food Preparation and Serving
- ☐ Office and Administrative Support
- ☐ Accounting and Finance
- ☐ Healthcare and Medical
- ☐ Industry and Manufacturing
- ☐ Law Enforcement
- ☐ Business Management
- ☐ Other



EXPERIMENT 2 - PROCEDURE



Scenario text

- With markup -

text = potential "who" **text** = potential "what" **text** = potential "where" **text** = potential "when"

The Gray and Teal groups do not employ suicide bombers.

The Turquoise group focuses on destroying energy infrastructure.

No attacks are being planned on religious organizations in Sigmaland.

The target is in a coalition country (Mulan, Xiland, Omicronland, Piland, or Sigmaland).

An attack is being planned for the first month of the year.

The Rose group may be involved.

There is a lot of activity involving the Rose group.

Embassies in Piland were recently attacked and evidence of more attacks has been found.

There are reports that spent nuclear fuel is missing in Muland.

The largest bank in Piland has 4 machine gun emplacements on its roof.

No traces of members from the Blue group have been found in Sigmaland.

Mulan is in the process of deploying more troops to protect its military bases.

The Blue and Turquoise groups use very small cells.

Army bases in Piland are increasing night patrols.

Members of the Silver, Turquoise, Teal, Sienna, and Rose groups have experience with explosives.

Reports from Omicronland, Piland, and Sigmaland indicate surveillance ongoing at public gathering places.

Attacking buildings when there are many people present increases casualties.

Time remaining: 19:41

Terrorist plot:

Who? -- Select group --

What? -- Select target --

Where? -- Select country --

When? -- Select month -- at -- Select time -- -- Select AM/PM --

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

- Without markup -

The Gray and Teal groups do not employ suicide bombers.

The Turquoise group focuses on destroying energy infrastructure.

No attacks are being planned on religious organizations in Sigmaland.

The target is in a coalition country (Mulan, Xiland, Omicronland, Piland, or Sigmaland).

An attack is being planned for the first month of the year.

The Rose group may be involved.

There is a lot of activity involving the Rose group.

Embassies in Piland were recently attacked and evidence of more attacks has been found.

There are reports that spent nuclear fuel is missing in Muland.

The largest bank in Piland has 4 machine gun emplacements on its roof.

No traces of members from the Blue group have been found in Sigmaland.

Mulan is in the process of deploying more troops to protect its military bases.

The Blue and Turquoise groups use very small cells.

Army bases in Piland are increasing night patrols.

Members of the Silver, Turquoise, Teal, Sienna, and Rose groups have experience with explosives.

Reports from Omicronland, Piland, and Sigmaland indicate surveillance ongoing at public gathering places.

Attacking buildings when there are many people present increases casualties.

There are fewer attacks in the heat of Summer (June and July).

The Silver group does not work in Piland.

Time remaining: 19:55

Terrorist plot:

Who? -- Select group --

What? -- Select target --

Where? -- Select country --

When? -- Select month -- at -- Select time -- -- Select AM/PM --

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



EXPERIMENT 2 - PROCEDURE



Workload (NASA-TLX) and preference

Mental Demand <i>How mentally demanding was the task?</i>	Very Low	Very High
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Physical Demand <i>How physically demanding was the task?</i>	Very Low	Very High
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Temporal Demand <i>How hurried or rushed was the pace of the task?</i>	Very Low	Very High
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Performance <i>How successful were you in accomplishing what you were asked to do?</i>	Perfect	Failure
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Effort <i>How hard did you have to work to accomplish your level of performance?</i>	Very Low	Very High
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	
Frustration <i>How insecure, discouraged, irritated, stressed, and annoyed were you?</i>	Very Low	Very High
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	

Finally, recall that in training you were shown examples of text both with and without computer-generated highlighting. While you were only given the opportunity to solve full scenarios under one of these conditions, we are interested to know which condition you would overall prefer to work with.

Preference <i>If given the choice, which version of the task would you prefer to work with?</i>	Definitely without highlighting	Equal	Definitely with highlighting
	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <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 - PROCEDURE

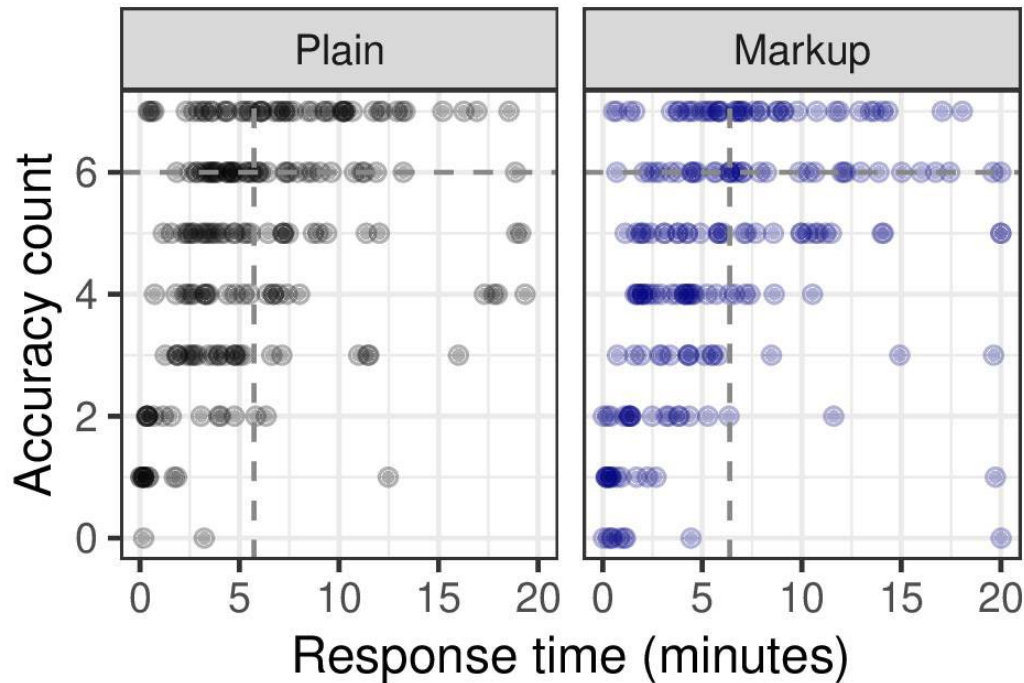


Trust in automation

	Disagree						Agree
	1	2	3	4	5	6	7
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>
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"/>



EXPERIMENT 2 - RESULTS



- **No significant difference in accuracy, speed**
 - Again, participants are more accurate when in preferred condition
- **Felt more successful with Markup**
- **Preferred Markup**



EXPERIMENT 2 - RESULTS



Demographics and correlation

- **Trust in automation ~ preference for markup:**
 - $r = 0.39$
- **Trust in automation ~ objective performance measures**
 - accuracy: $r = 0.06$
 - response time: $r = -0.05$

Question	Response category	N (All)	N (Filtered)
Language	No	27	16
	Yes	172	133
Gender	Female	86	66
	Male	112	82
	Other/prefer not to say	1	0
Age	18-29	78	45
	30-49	91	78
	50-64	25	22
	65+	5	4
Education	High school graduate or less	20	15
	Some college	70	55
	College degree or more	109	79
Occupation	Science and technology	67	56
	Arts, entertainment, and media	9	7
	Education	7	5
	Legal	13	13
	Sales	17	10
	Food preparation and serving	12	6
	Office administration and support	3	3
	Accounting and finance	20	17
	Healthcare and medical	7	2
	Industry and manufacturing	18	11
	Law enforcement	15	9
	Business management	11	10
	Other	0	0

Correlation coefficients

Covariates		Plain	Markup
Language	Accuracy	0.31	0
	Response time	-0.31	-0.31
	Preference	-0.08	-0.06
Gender	Accuracy	-0.1	-0.1
	Response time	0	-0.14
	Preference	0.05	0.01
Age	Accuracy	0.05	0.03
	Response time	0.03	0.14
	Preference	-0.22	-0.13
Education	Accuracy	-0.05	-0.2
	Response time	0.06	0.1
	Preference	-0.05	-0.18
Occupation	Accuracy	0.12	0
	Response time	0.21	0.06
	Preference	0.04	0.01



DISCUSSION



- Results summary
 - Experiment 1 - Instead of helping, markup hurt performance and was dispreferred to plain text.
 - Experiment 2 – Even with “ideal” markup, no difference in performance, but markup preferred.
- Warning to researchers trusting that automated IE will be helpful to end user
- Objective performance vs. Subjective experience
 - Even if markup does lead to better performance, it may improve/hurt experience
- Trust in automation
 - Correlation with preference
- Filtering
 - These were responses that were too quick to represent true attempts to read the texts and identify the hypothetical adversary attack.
 - These participants were roughly twice as likely to report that English was not their native language, often provided incoherent free-text strategy descriptions.
 - Are Turkers generally not willing to put in the work necessary to do well at this task?
 - This population does not necessarily predict the performance of any other population, important to include the specific intended end user in the testing cycle.



THANKS!



- **Coauthors**
 - Justine Caylor
 - Michelle Vanni
 - Sue Kase



PREFERENCE



Participants perform qualitatively more accurately in their preferred condition, though Plain is always slower.

Experiment 1

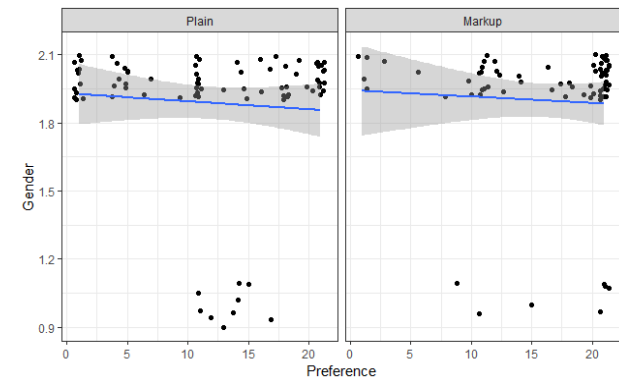
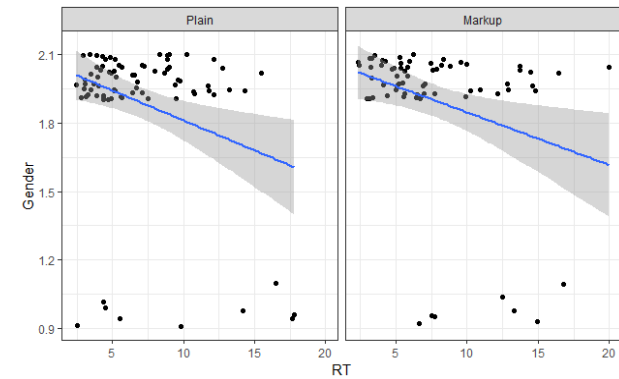
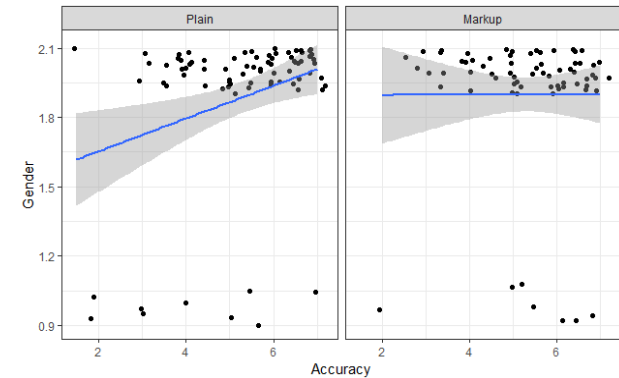
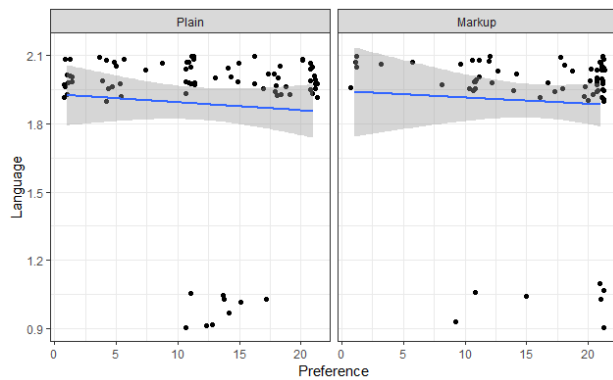
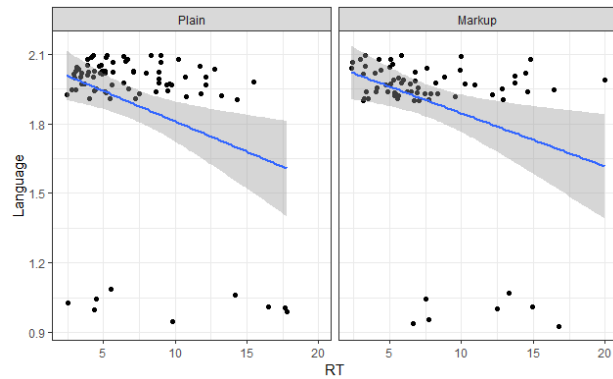
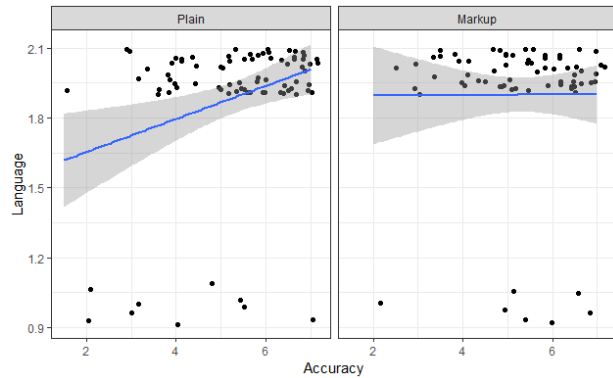
Preference	N	Condition	Median accuracy count	Median response time (min)
Plain	66	Plain	6	6.82
		Markup	4	7.15
Markup	30	Plain	4	4.10
		Markup	5	5.25

Experiment 2

Preference	N	Condition	Median accuracy count	Median response time (min)
Plain	70	Plain	6	7.00
		Markup	5	8.11
Markup	178	Plain	5	4.89
		Markup	6	6.39

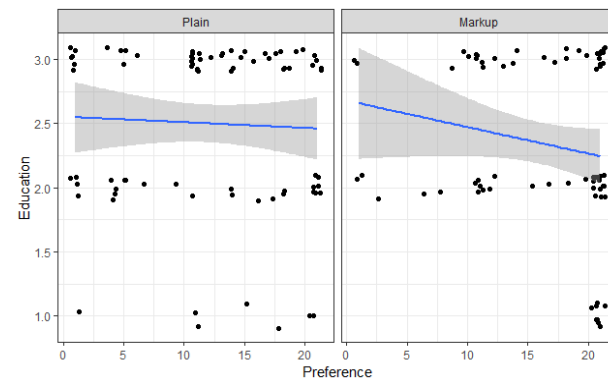
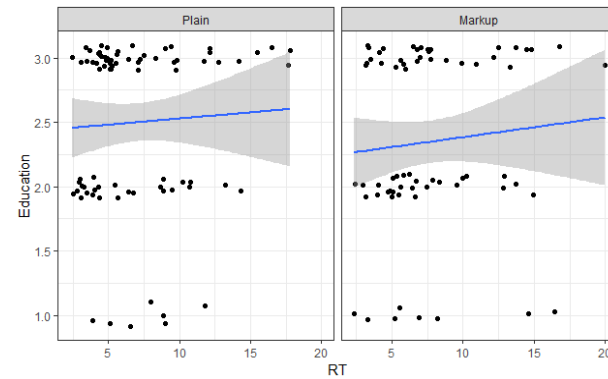
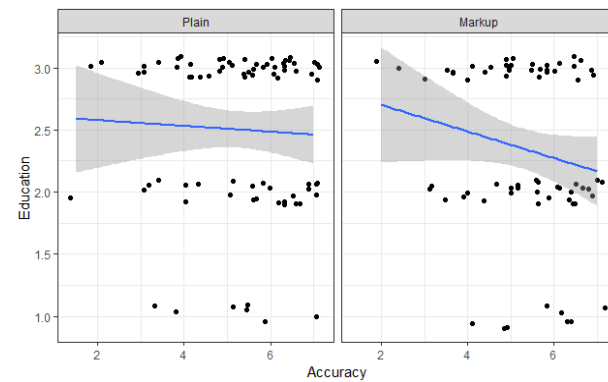
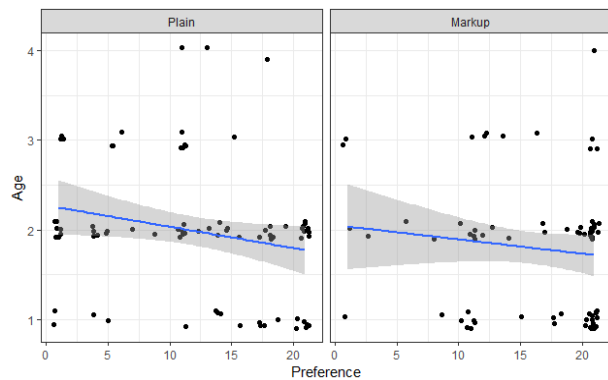
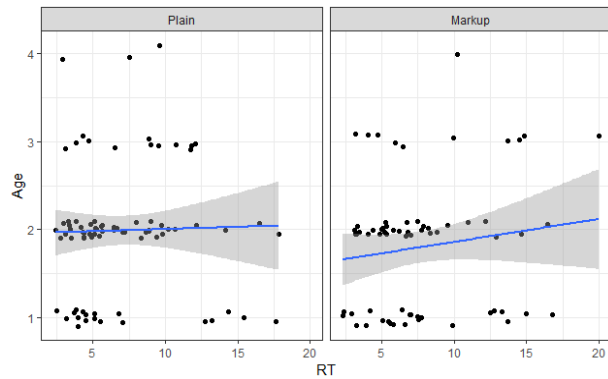
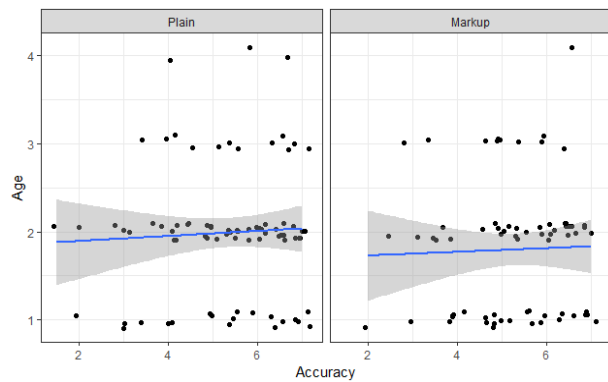


CORRELATIONS





CORRELATIONS





CORRELATIONS

