Common ground in collaborative intelligence analysis: an empirical study

Proceedings British HCI 2016 – Fusion, Bournemouth, UK. — 13th July 2016

Dr. Sean Xavier Laurence

— The Joint intelligence Training Group,
Royal School of Military Survey, THATCHAM, UK

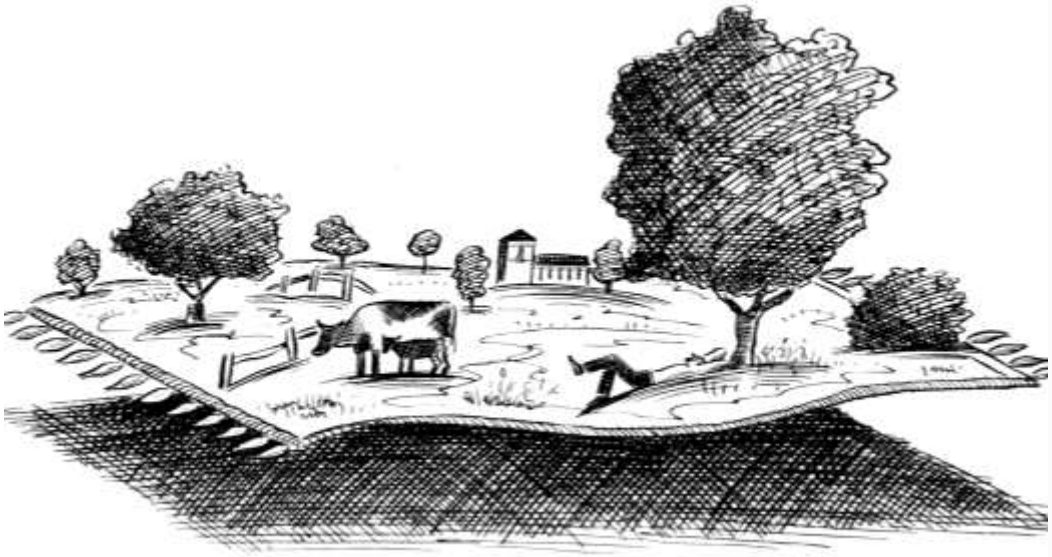
Dr. Simon Attfield

Dr. Bob Fields

(— Middlesex University, London, UK)

Common ground
| Conversational grounding

- **So what is Common Ground?**

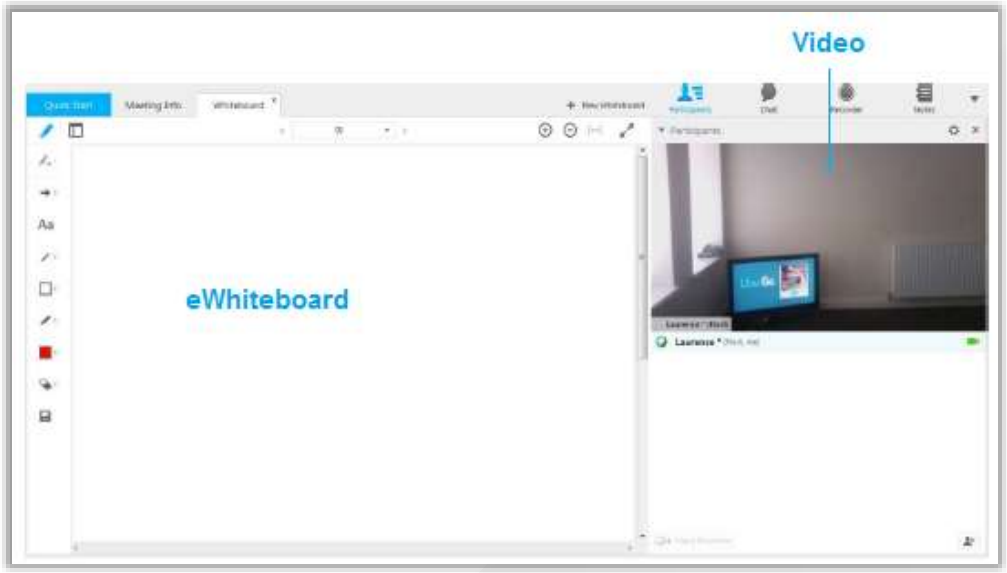


So a bit of Background info

— Intelligence analysis is much more than an individual activity.

— Using video offers a way to replicate non-verbal communicative cues in F2F communication.

— A shared visual workspace offers a common reference place to explore task artefacts interactively.



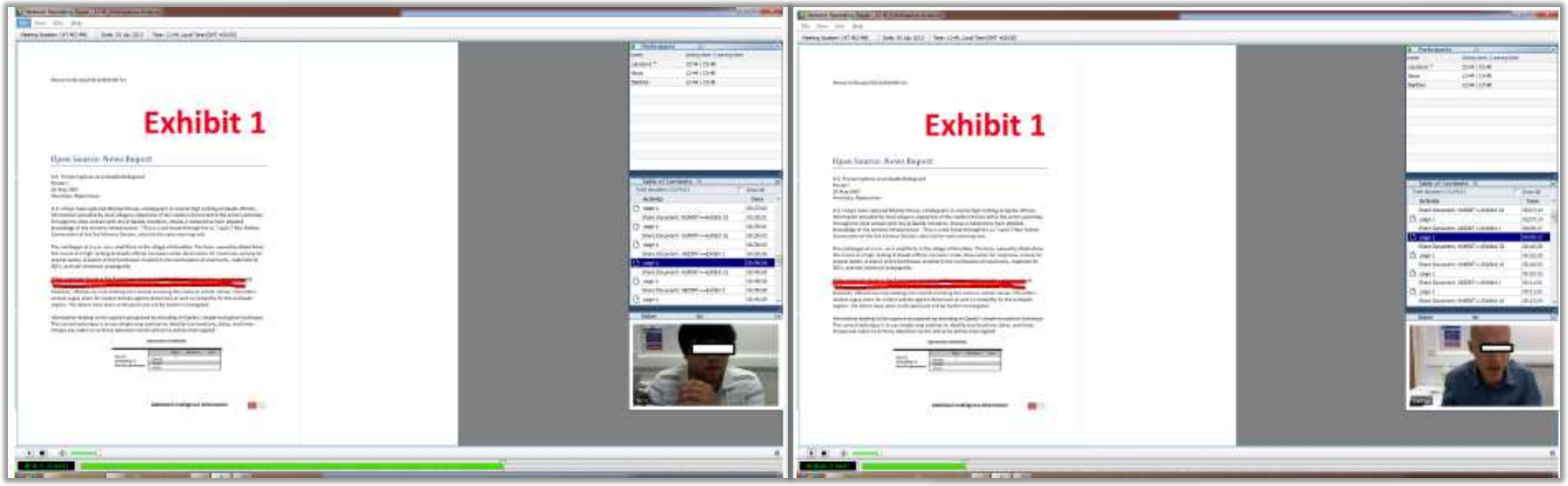
— An electronic bounded space with support for reciprocity, and for sharing, visualising, and interacting with task artefacts in real time.

—Laurence 2015

Motivation— What the study clarifies

- Prior use/impact studies of collaboration technology have focused on tasks that require physical handling.

Identical views of a shared visual workspace



Research Design

- 2x2 between-subject factorial design.
- Two independent variables — presence/absence of shared visual workspace and video;
- One dependent variable —conversational grounding effort.

- Two separate labs.
- Environment — iMac desktop units running the Cisco WebEx video conferencing software with integrated VoIP.

- Convenience sampling
- 56 participants;
- 2-member teams | 28 teams;
- 25 undergrads | 31 post-grad;
- 18 -48 age range.

Hypotheses

H1. Teams using shared visual workspace will construct repair-episodes at a lower rate than those without.

H2. Teams using video will construct repair episodes at a lower rate than those without.

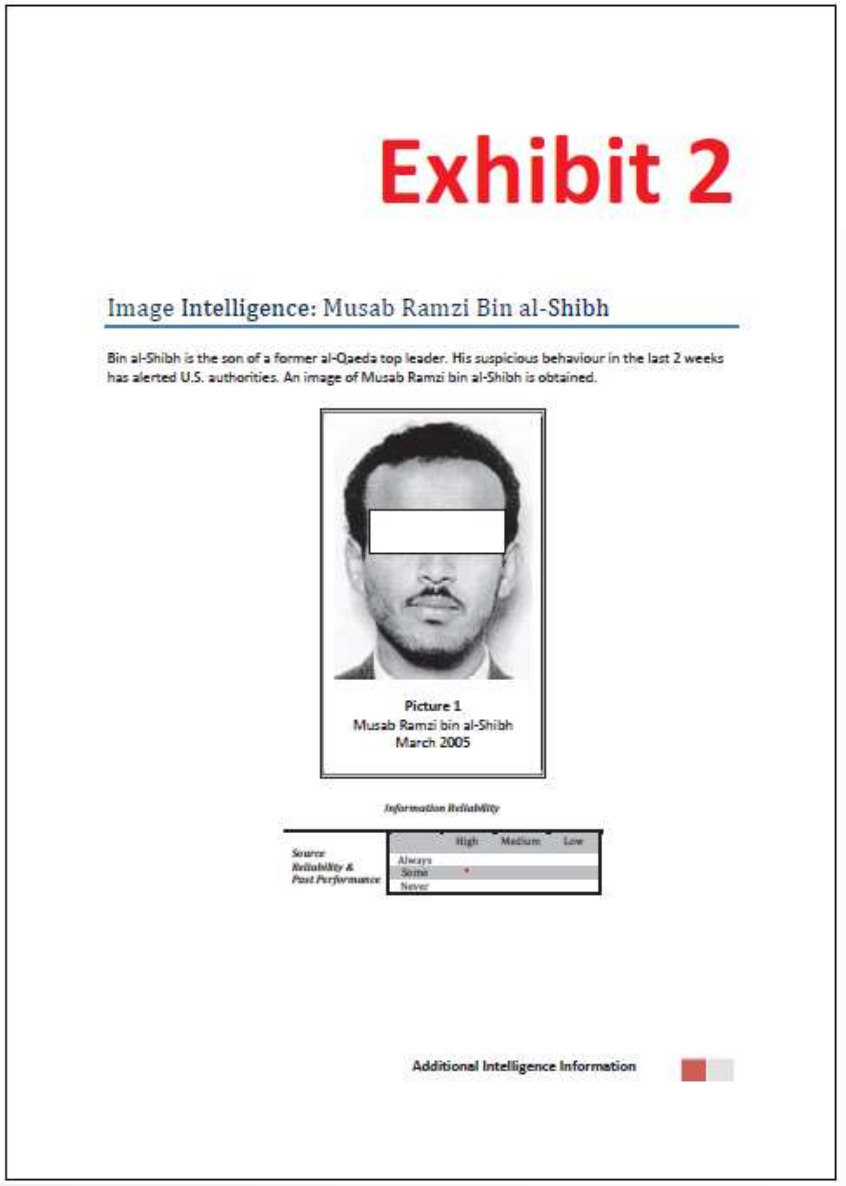
H3. An association exists between shared visual workspace and video in terms of the rate of repair-episodes.

H4. The rate of repair-episodes will decrease as the task progresses.

— The Intelligence Analysis Tasks

Task 1: Assess if Farah has allegiance with a new terrorist cell network recently formed in his home town of Dissibad.

Task 2: Identify the five most plausible events or evidence from the intelligence data which supports any conclusion reached above.



Data collection | analysis — Coding schema

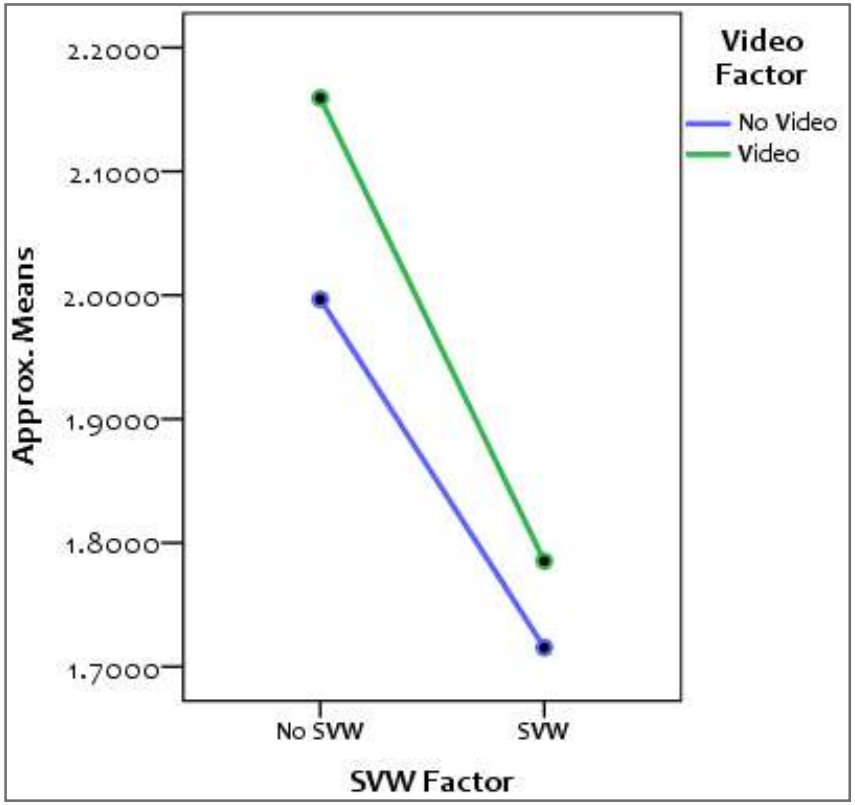
To access the full codebook visit: http://eprints.mdx.ac.uk
— (Laurence, 2015)

Codes	Code Description
	Some examples: *Should we go over Exhibit 52 — Um, just forget it.*
4. Understanding Check	(1). To code any utterance a person produces to check 'self' or the message recipient's understanding of a previous utterance; or current state of understanding with respect to the task.
Subcodes	
PRIOR ACTION-CHECK	(2). To code questions or phrases for checking whether a previous 'action' has been understood by another group member. **Some examples:** *Do you see what I'm doing?; Did you see what I just shared?*

Coding schema — more excerpts

Codes	Code Description
1. Introduce-Dialogue	To code a new discourse unit — a new point, idea, topic, or information. **Annotation Rules** Excludes responses or utterances evoked or elicited by a person's conversation partner.
5. Repair-Request	(1). To code a message recipient's requests for the speaker to repeat, rephrase or simplify an utterance. **Some examples:** *Sorry?; Excuse me?; Pardon?; I don't understand; Say that again please* *A1: Did Kris finally resit his Mth 401 exams?* *A2: Hu:h?* **(*Repair-Request)** *A1: Did Kris retake his Mth 401 exams afterwards?* **(*Repair)** *A2: I have no idea*

— Data analysis and results



Estimated marginal means for repair episodes

2-Way ANOVA Results

H1 | main effect of SVW

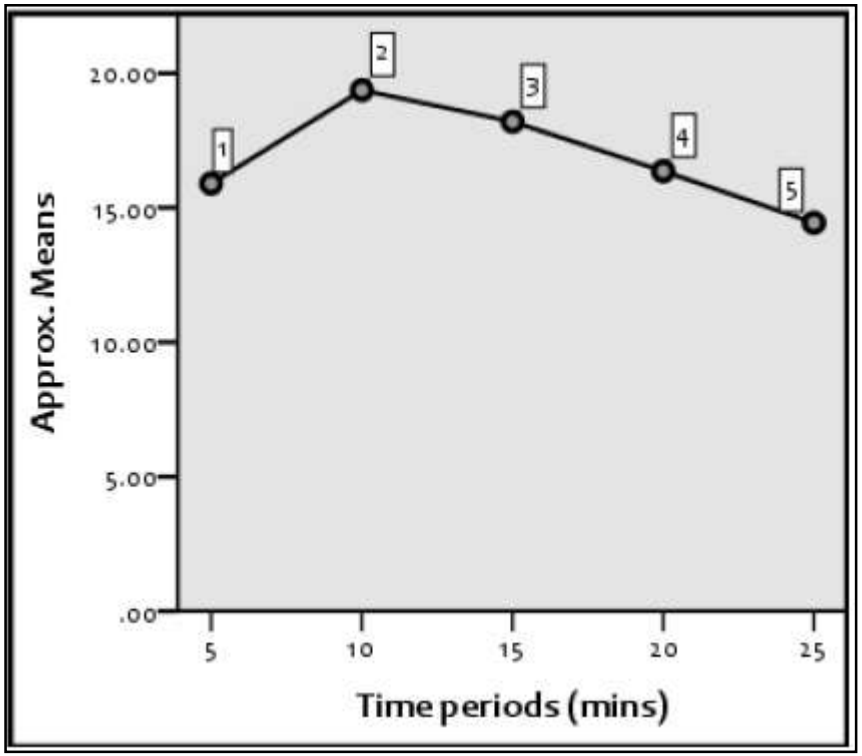
($F(1,24) = 4.988$, $\rho=.035$, $\eta^2 = .172$).

H1 is supported

No main effect for video; and no interaction effect for the two independent variables.

H2 & H3 not supported

Other results — Pearson's correlation



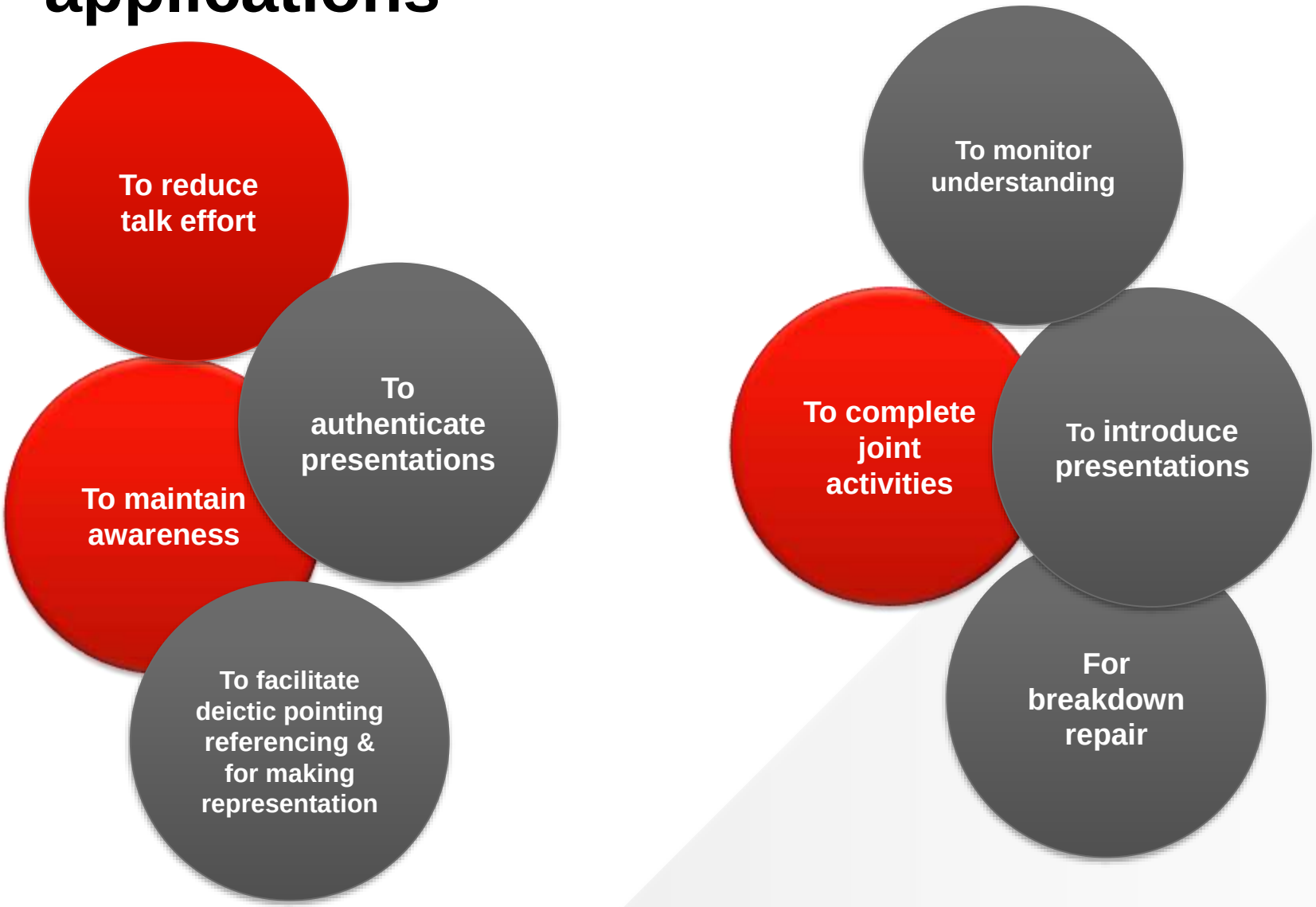
Rate of repair episodes over time

Measuring the **rate of repair-episodes** against **time period** in the experiment.

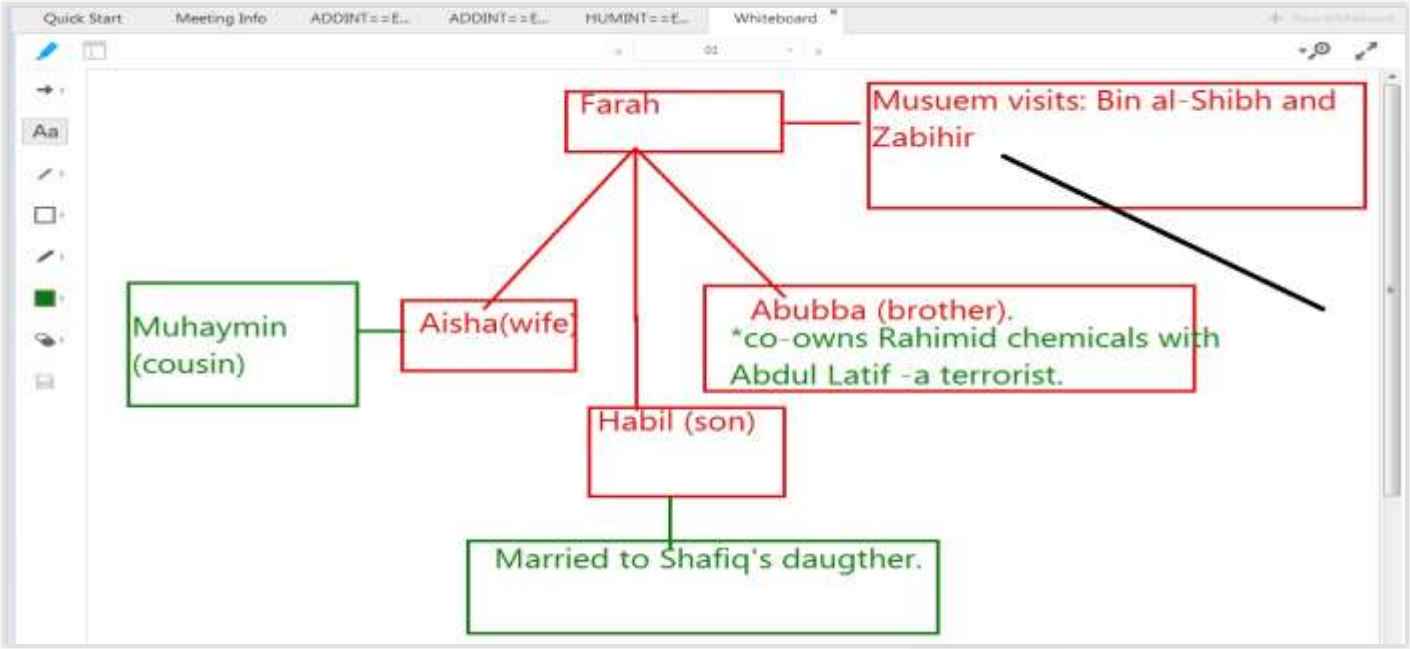
- At T1, the rate of repair-episodes was higher; T1 was excluded from the Pearson's correlation test.
- At T2 –T5, rate of repair-episodes progressively declines as task progresses.
- The relationship between the rate of repair-episodes and time period participants spent working on the task was statistically significant. ($\rho < .01$; $\rho = -.994$; 2-tailed).

H4 is supported

—8 SVW applications



To maintain awareness



Dialogue 1

Bob: Erm, that can't be right! I think what you meant to say there is Zarqawi and not Zabihir. Zabihir is the name of the town! What we said was that Bin alShibh and Zarqawi both frequented the museum regularly. You should change it.

Alex: OoH yes yes!

To minimise communicative effort

Dialogue 2

Alex: Have you come across someone named Rana Ba-seyl-rhat yet?
Bob: God! These names!
Alex: Let me see if I can write on this thing.
Bob: What was the name you said again?
Alex: Just look on the board, I'll show you there.

(He starts writing on the SVW;
he draws a "square" shape over the text)

Bob: Aaargh! Okay! — "Rana Baseerat".

To facilitate, introduce & complete a presentation.

Dialogue 3

Bob: Hang on a minute. Actually before you take this off the screen; — this chap here on this ITC statement — Abdul Amar Qazafi.
Alex: What about him?
Bob: Dude owns a farm where a terrorist bodyguard was arrested in 2007.
Alex: What dude? What farm?
Bob: "Abdul Amar " —number 7 on this ITC list you have up. One of my records here, he owns a farm that was raided by the counterterrorism team.
Alex: Hmm! Okay! Interesting!

Discussion

Our results are explained by …

- Differences in communication media affordances.
- Attenuation differences in media communication cues, visual cues and awareness nuances, etc.
- A collaborative framework with fewer visual and communication cues generally increases the effort towards grounding — vice versa.

Conclusion

Drawbacks:

- Catch 22s — Intelligence domain studies issues.
- Possible existing affiliations between group members.
- Making broad generalisation from a student sample.

Future research:

- Explore the performance differences of the intelligence tasks.
- Explore if a positive correlation exists between **prior experience** with **svw** and **rate of repair-episodes**.

Q/A

Arguments for creating new model

- No one model fits all.
- Substantial differences in code categorizations and their definitions.
- Codes intended for use with specific units of analysis in mind
- Codes created for use with specific data types E.g., threaded discussions, chat, etc
- Contexts and tasks types for which they were designed may not be suited.
- Often contain numerous ambiguities which it makes it harder to distinguish them.
- Often lacked classifications that can be quantifiable and compared across media conditions.

Pros for our coding schema

- Implemented strong annotation rules for our coding schemas.
- Independent raters obtained homogenous results, which demonstrating that their understanding and interpretation of the annotation rules and codes were similar.