Structure as an Aid to Good Science

Roy Maxion

Dependable Systems Laboratory
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
Email: maxion@cs.cmu.edu

22-26 January 2015

IFIP Working Group 10.4
Workshop on The Science of Cyber Security
Tortworth, Bristol, UK

Today's goal

- To sensitize you to a few simple things ... and how they could influence the science of security:
  - **Story grammars/schemas** – pieces of a story (paper), in unexpected order, can cast a claim into doubt or certitude.
  - **Structured abstracts** – enable readers to grasp quickly and completely the essence of a paper.
  - **Structured reporting** – the way a paper is structured will influence authors and readers alike.
  - **Structured reviews** – the way a review form is structured will influence the referees, either positively or negatively.
- Initiate discussion; improve science; shift the culture.

Story grammar (script, schema) [SchankAbelson77]

- A word about how important structure is in our daily lives …
- A story grammar/script helps readers understand the structure of a text.
- At an elementary level, a story grammar may include such elements as the characters, the setting, and a beginning, middle and end … setting the expectations of the reader (or listener).

Story grammar / script

- Essentially, a story grammar (or script) – just like any other grammar – tells the "parser" what to expect.
- In a "restaurant" script, for example, you wouldn't expect the food to arrive before the order has been placed.
- A grammar provides structure, making it easier to "parse" the story, and to comprehend it.
- Structure for how we prepare abstracts, papers and reviews is equally important, but seldom seen – and this makes life harder than it needs to be.
- Plus … when everyone violates the grammar, it becomes normal not to expect much structure, and without that structure or expectation, errors of commission and omission can creep in … which is hard to detect, because these errors may not violate expectations (which were never established in the first place).

Story grammars pertain to all "stories"

- Abstracts
- Papers
- Reviews
- Proposals

Violating a reader's expectations can have unfortunate consequences … and often does.

What is an abstract?

- An abstract is a self-contained, short statement that describes a larger work.
- Two general types:
  - **Descriptive** – a rough outline of the paper; usually very short (~100-200 words); often a come-on to get you to read the paper – what we usually see.
  - **Informative** – acts as a surrogate for the work; usually longer (~500 words); contains quantitative details (e.g., results); enables prospective readers to judge the relevance of the work; not what we usually see. Has specific structure. Sometimes called "**structured**."

Issues in typical abstracts

- Hard to write
- Often hard to understand
- Frequently just an advertisement to get you to download the paper
  - … only to find out the paper's results, which are usually omitted from the abstract;
  - and, all too often, to discover that the paper is about something different than the abstract suggested.
- Often don't convey the critical details.
- Difficult to use for PC paper bidding.
- Hard or impossible to scrape.

What is a structured abstract?

- In their most basic form, structured abstracts are organized in accordance with distinct, labeled sections:
  - Objective
  - Methods
  - Results
  - Conclusions
- These headings are mutable, and appear in bold type (in the abstract itself) to set them off and to make them easy to read/scan.
- Purpose: to facilitate rapid comprehension.

Nine starting components

[MostellerEtAl04]

1. Background and context
2. Purpose / objective / research question
3. Setting – conditions under which the work was done
4. Population / participants / subjects
5. Intervention / program / practice
6. Research design
7. Data collection and analysis
8. Findings / results
9. Conclusions / recommendations

Can choose from amongst these, as appropriate for the forum.

Wouldn't it be nice to get all this out of an abstract?

Structured abstract – the essentials

- Background
- Aim
- Data
- Method
- Results
- Conclusions

An example ...

Free vs. Transcribed Text for
Keystroke-Dynamics Evaluations

Example

- **Background.** One application of keystroke dynamics is continuous re-authentication: confirming a typist's identity during normal computer usage without interrupting the user.

Example

- **Aim.** In laboratory evaluations, subjects are typically given transcription tasks rather than free composition (e.g., copying rather than composing text), because transcription is easier for subjects. This work establishes whether free and transcribed text produce equivalent evaluation results.

Example

- **Data.** Key-hold and keydown-keydown times were derived from 40 samples of free-text and transcription typing (~125 characters each) from 20 subjects.

Example

- **Method.** Twenty subjects completed comparable transcription and free-composition tasks; two keystroke-dynamics classifiers were implemented; each classifier was evaluated using both the free-composition and transcription samples.

Example

- **Results.** Transcription hold and keydown-keydown times are 2–3 milliseconds slower than free-text features; t-tests showed these effects to be significant. However, these effects did not significantly change evaluation results.

Example

- **Conclusions.** The additional difficulty of collecting freely composed text from subjects seems unnecessary; researchers are encouraged to continue using transcription tasks.

Structured abstract (paper on a slide)

- **Background.** One revolutionary application of keystroke dynamics is continuous re-authentication: confirming a typist's identity during normal computer usage without interrupting the user.
- **Aim.** In laboratory evaluations, subjects are typically given transcription tasks rather than free composition (e.g., copying rather than composing text), because transcription is easier for subjects. This work establishes whether free and transcribed text produce equivalent evaluation results.
- **Data.** Key-hold and keydown-keydown times were derived from 40 samples of free-text and transcription typing (~125 characters each) from 20 subjects.
- **Method.** Twenty subjects completed comparable transcription and free-composition tasks; two keystroke-dynamics classifiers were implemented; each classifier was evaluated using both the free-composition and transcription samples.
- **Results.** Transcription hold and keydown-keydown times are 2–3 milliseconds slower than free-text features; t-tests showed these effects to be significant. However, these effects did not significantly change evaluation results.
- **Conclusions.** The additional difficulty of collecting freely composed text from subjects seems unnecessary; researchers are encouraged to continue using transcription tasks.

Adoption of structured abstracts by medical journals

- **Background:** The use of a structured abstract has been recommended in reporting medical literature to quickly convey necessary information to editors and readers. The use of structured abstracts increased during the mid-1990s; however, recent practice has yet to be analyzed.
- **Objectives:** This article explored actual reporting patterns of abstracts recently published in selected medical journals and examined what these journals required of abstracts (structured or otherwise and, if structured, which format).
- **Methods:** The top thirty journals according to impact factors noted in the "Medicine, General and Internal" category of the ISI Journal Citation Reports (2000) were sampled. Articles of original contributions published by each journal in January 2001 were examined. Cluster analysis was performed to classify the patterns of structured abstracts objectively. Journals' instructions to authors for writing an article abstract were also examined.
- **Results:** Among 304 original articles that included abstracts, 188 (61.8%) had structured and 116 (38.2%) had unstructured abstracts. One hundred twenty-five (66.5%) of the abstracts used the introduction, methods, results, and discussion (IMRAD) format, and 63 (33.5%) used the 8-heading format proposed by Haynes et al. Twenty-one journals requested structured abstracts in their instructions to authors; 8 journals requested the 8-heading format; and 1 journal requested it only for intervention studies.
- **Conclusions:** Even in recent years, not all abstracts of original articles are structured. The eight-heading format was neither commonly used in actual reporting patterns nor noted in journal instructions to authors.

[Journal of the Medical Library Association, Nakayama et al. 2005]

Science Communication

- **Background**. Previous research with structured abstracts has taken place in mainly medical contexts. This research indicated that such abstracts are more informative, more readable, and more appreciated by readers than are traditional abstracts.
- **Aim**. The purpose of this study was to test the hypothesis that structured abstracts might also be appropriate for a particular psychology journal.
- **Method**. 24 traditional abstracts from the Journal of Educational Psychology were re-written in a structured form. Measures of word length, information content and readability were made for both sets of abstracts, and 48 authors rated their clarity.
- **Results**. The structured abstracts were significantly longer than the original ones, but they were also significantly more informative and readable, and judged significantly clearer by these academic authors.
- **Conclusions**. These findings support the notion that structured abstracts could be profitably introduced into many journals.

J. Hartley, *Science Communication*, Vol. 24, No. 3, March 2003, pp. 366-379

The case in favor …

- Structured abstracts …
  - contain more information
  - are easier to read
  - are easier to search
  - are easier to recall
  - facilitate peer review
  - are easier to write
  - are usually more consistent and clear
  - are more easily evaluated when there are many
  - are more easily comprehended due to readers being able to isolate sections, and read about key aspects of the paper
  - easier to decide to spend $25-35 to download (or not)
  - easier to scrape; perform meta-analyses

Structured abstract – pros and cons

- Pros
  - Organizes one's thoughts
  - Easy to comprehend the entire study at a glance
  - No mystery stories
  - Easy to decide if you want to read the entire paper (you might want to, to see the methods)
  - Amenable to meta-analysis
- Cons
  - The opposite of everything in the Pro column
  - Harder to write (but not after you try it a couple of times)
  - It's easier not to do it.
  - Longer than 150 words.

Consider ...

- When looking through 100-200 abstracts, prior to choosing which papers you want to review, would structured abstracts be more helpful in making your decisions than the descriptive abstracts in use now?
- Helps readers, too; not just reviewers.
- Might be worth a try.

Uses for structured abstracts

- Certain types of readers find structured abstracts particularly beneficial:
  - Those who will not read an article in its entirety, but need to know the key facts; e.g., executives, principal investigators, and people trying to keep abreast of the field.
  - Those who have previously read the article in its entirety, and need to recall key findings without having to re-read the article, such as researchers conducting a systematic review of the literature.
  - Those who are trying to determine whether or not to read a particular article.

Structured abstracts are in many journals

- Because they facilitate reading and retaining information from articles, many peer-reviewed journals are adopting structured abstracts as their preferred format for abstracts, including
  - IEEE Transactions on Professional Communication
  - Which provides guidelines for various kinds of studies
    - Research article
    - Case study
    - Tutorial
    - Literature review
    - Teaching case

Structured abstracts - summary

- Easier to read and to write
- Usually clearer than text/descriptive abstracts
- Start using structured abstracts now
- At least require them for paper submissions, if not for the papers themselves
- Scrape them to do literature reviews and meta-analyses
- Be the leader; be (among) the first to adopt

Issues in papers

- Lack of structure …
  - Impedes comprehension (no “story”)
  - Creates opportunity for errors of omission
    - failure to ensure validity, reproducibility, etc.
    - failure to describe apparatus, materials, subjects, etc.
    - failure to conduct power analyses
    - failure to conduct representative sampling
    - failure to state the experimental design
    - failure to justify every decision
    - … and a host of similar sins

Justifying all decisions … means what?

- Study of typing and stress …
  - Three conditions: baseline, stress, baseline
  - Induce baseline, then stress, then return to baseline
- How do you know that a stressor will do its job?
  - Use vetted stressor from the literature.
- How do you know that the subject is stressed?
  - Independent physiological measures
- What strings will subjects type?
  - Easy to type, 30 characters
- How to know that the strings are easy to type?
  - Mechanical Turk study
- Will the strings themselves evoke a stress emotion? How do you know?
  - International Affective Picture System
- No stone left unturned … harder without structure

Why focus on experimental science?

The principle of science, the definition, almost, is the following: the test of all knowledge is experiment.

Experiment is the sole judge of scientific "truth".

But what is the source of knowledge? Where do the laws that are to be tested come from?

Experiment, itself, helps to produce these laws, in the sense that it gives us hints.

- Richard Feynman, Lecture on Physics

Some hallmarks of a good experiment

- Valid
- Reliable / Repeatable
- Reproducible
- Properly / transparently reported

Hallmarks of a good experiment (1/4)

- Valid
  - Internal - An experiment is internally valid if there are no alternative explanations for the outcome, other than the one posited for the experiment.
    - Example – distinguishing users by mouse movements … but letting users choose their own web content means that the content could have explained the outcome, not the user mousing style
  - External - An experiment is externally valid if the conclusions drawn from the experiment can be extended beyond the bounds of the experiment.
    - Example – data from social media study were biased; sites like Facebook and Twitter attract different demographics, so conclusions based on one site do not generalize to the broader population. (Science, 112814)

Hallmarks of a good experiment (2/4)

- Reliable/repeatable
  - Repeatability refers to the variation in repeated measurements taken by a single person or instrument … on the same item … and under the same conditions; we seek high agreement from one measured instance to another.
  - A measurement is said to be repeatable when this variation in repeated measurements is smaller than some agreed limit.

Hallmarks of a good experiment (3/4)

- Reproducible
  - Measurements give the same results each time they are taken, irrespective of who does the measuring.
  - Reproducibility relates to the agreement of experimental results with independent researchers using similar but physically different test apparatus, and different laboratory locations, but trying to achieve the same outcome as was published in an article.
  - Replication or reproduction allows an assessment of the control on the operating conditions, i.e., the ability to reset the conditions to some desired value. Ultimately, replication reflects how well the procedure was operationalized.

Hallmarks of a good experiment (4/4)

- Properly / transparently reported
  - All methodological details are provided (preferably together, in one place), enabling readers to reproduce the experiment (even if only mentally), and to obtain the same results.
  - The details of the experimental method are what enable readers to judge the validity of the work. When we leave out these details, it cripples the reader, although only some readers notice.

Parts (grammar) of an experimental paper

- Title
- Author(s)
- Abstract
- Introduction
- Problem being solved
- Background and related work
- Approach
- **Method** →
  - Apparatus & instrumentation
  - Materials
  - Subjects / objects
  - Instructions to subjects
  - Design
  - Procedure
- Data
- Analysis
- Results
- Discussion
- Limitations
- Conclusion
- Future
- Acknowledgements
- References
- Appendices
- Endnotes and footnotes

Parts of an experimental paper

- These are the sections from which the details emerge that bear on experimental validity and reproducibility.
  - Apparatus & instrumentation
  - Materials
  - Subjects / objects
  - Instructions to subjects
  - Design
  - Procedure

Aphorismus ...

The unity of all science consists alone in its method, not in its material.

Karl Pearson
The Grammar of Science
Meridian Books: NY. 1911
(First published in 1892)

Let's briefly examine these elements

- Method
  - Tells the reader how the experiment was conducted, providing enough information so that readers can verify and reproduce the results.
- Sections
  - Apparatus & instrumentation
  - Materials
  - Subjects / objects
  - Instructions to subjects
  - Design
  - Procedure
  - Analysis
- There are others, but we'll focus on these.

Methodological elements - apparatus

- Apparatus & instrumentation
  - A general description of the apparatus, including any details that might affect the outcome of the experiment; name and model number of equipment, and last calibration date.
  - Hardware
  - Software
  - Packages (e.g., R for statistics)
  - Data sets
  - Any “infrastructure” used

Methodological elements - materials

- Materials
  - Describe the stimulus materials in enough detail so that a reader could generate the same or equivalent stimuli.

Methodological elements - subjects

- Subjects / objects
  - Describe the total number of subjects/objects, the number receiving each treatment, the population from which the subjects/objects were drawn, the circumstances under which they were selected (sampling scheme).
  - The nature of the experiment will determine what other attributes are relevant.

Methodological elements - instructions

- Instructions to subjects
  - Describe what the subjects (if used) were asked to do; verbatim if possible.

Methodological elements - design

- Design
  - Include a description of the dependent and independent variables (and any other features being measured), how the experimental and control groups were constituted, and how subjects were assigned to these groups (e.g., at random).
  - Provide the common name of the analytical design used (case study, case/control, cross-sectional, randomized controlled trial, etc.).

Aside: levels/quality of evidence

Each study design provides a certain level of evidence to support a claim. In order, from bad to better …

- No evaluation at all
- Editorials, “expert” opinions
- Weakly supported claims
- Case series, case reports
- Cross-sectional studies
- Case-control studies
- Cohort studies
- Randomized controlled trials
- Systematic reviews (meta-analyses)

Methodological elements - analysis

- Analysis
  - Provide descriptive statistics summarizing the data in a readily comprehensible form, and inferential statistics which test the likelihood that the obtained results were not due to chance.

Methodological maxim

You may forget a few details …
but they won't forget you.

Paper structure – summary / counsel

- Structure helps in both writing and reading.
- Work as a group (e.g., 10.4) to define what constitutes a structure sufficiently complete to ensure good reporting and good science.
- Implement that structure, with guidelines, for your publication venues (e.g., DSN).

Issues in reviews

- Variability / consistency
  - NIPS study, December 2014
  - Between two parallel PCs, 57% of the papers accepted by the first committee were rejected by the second one, and vice versa – arbitrariness?
  - Phenomenon probably not confined to NIPS
- Completeness, coverage of the right details
- Emphasis on novelty (misconstrued as cool)
- Many referees can't remember all the things to evaluate (can happen to anyone)
- Broader impacts? Intellectual merit?

Review criteria

- If the structure of papers is to change, then reviewing guidelines and forms must also change.
  - And perhaps ought to change anyway.
- More-complete review forms would help
  - Like a checklist
  - Adds structure
  - Keeps important things from being overlooked

Reviewer's checklist (excerpt)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomi			
Seque generati	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	

Reviewer's checklist (I)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance hover mouse here)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	

Reviewer's checklist (II)

Section/Topic	Item No	Checklist item	Reported on page No
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample siz	7a	How the sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	

LASER review form; excerpts

- Problem statement:
  - Is there a clear statement of the problem being solved, or the hypothesis being tested, or the research question being asked?
- Evidence:
  - Are the claims/findings supported with methodologically competent/sufficient evidence?
- Contribution/science:
  - In what ways does this paper advance the field? Does the science represent a meaningful contribution to the literature?
- Reproducibility:
  - Is the experimental work and methodology sufficiently well described to allow others to reproduce the experiment?

Other review forms/checklists

- CONSORT - checklist, random controlled trials
- ENTREQ - biomedical research
- COREQ - consolidated criteria for reporting qualitative research
- SPIRIT - clinical trial protocols
- CARE - case-report critical assessment
- SQUIRE - standards for quality improvement reporting excellence
- STARD - studies for diagnostic accuracy
- PRISMA - reporting systematic reviews

Structure / order in reviews

- Structure will help to …
  - Contain variability in reviews
  - Obtain better technical coverage in reviews
  - Evaluate papers using a more uniform and complete set of criteria
    - Leaves less to individual discretion
    - Still allows disagreement on common criteria

Reprise: now what?

- **Story grammar/schema** – "Stories" (abstracts, papers, reviews) are easier to read and comprehend when all the pieces are in the expected order.
- **Structured abstract** – a "bulleted" list summarizing the paper: e.g., background, aim, data, method, result, conclusion … easier to comprehend and to scrape.
  - Require in paper submissions, at least; also the paper itself.
- **Structured reporting** – a paper should follow a relatively standard order that puts the requisite details in their expected and proper places; encourages better scientific coverage, thereby improving the science.
  - Require a uniform format in papers, appropriate for domain.
- **Structured review** – a referee report that requires attention to specific and well-defined sections in a paper
  - Change the referee forms; make them public.
- Start a discussion; make small changes; improve the science; shift the culture; do it now.

Wrapping up

- Will structure solve all of our problems?
- Perhaps not, but it is a relatively easy, quick and effective approach that will go a long way toward improving our science.
- We can work together to improve our current and future science, in much the same way that 10.4 worked to improve the science of fault tolerance and reliability … with kudos to Al, Brian, Jean-Claude, et al.

Structure – abstracts, papers, reviews

I came, I saw, I was confused.

-- Breyten Breytenbach,
Return to Paradise (1993)

But perhaps no more … if we seize and nurture the opportunity to work together to improve our science, using structure and any other tools at our disposal.

Thank you ...

- Contact information
  - Roy.Maxion@cs.cmu.edu
  - Carnegie Mellon University
  - Computer Science and Machine Learning
  - +1-412-268-7556