

Institute for Data-Intensive Research

The Summer of V's

Volume, Velocity, Variety, Veracity

Fourth meeting



University of
St Andrews

21 September 2015

#SummerOfVs

Getting there

The meeting is in the Auditorium and Foyer of the **Byre Theatre**, University of St Andrews.

For directions, see:

<http://www.idir.st-andrews.ac.uk/vs>

12:30–1:20 pm

Buffet lunch (free).

1:20 pm

Professor AARON QUIGLEY, School of Computer Science, University of St Andrews (aquigley@st-andrews.ac.uk)

“Welcome to IDIR”.

1:25 pm

Dr DANIEL BARKER, School of Biology, University of St Andrews (db60@st-andrews.ac.uk)

“Welcome to the Summer of V’s”.

1:30 pm

Invited speaker:

Dr AMIT PUNDIK, Tel Aviv University (<https://en-law.tau.ac.il/profile/amitp> / amitp@post.tau.ac.il)

“Freedom and generalization: should criminal behaviour base-rates be admissible as evidence in court?”

See hand-out at the end of this programme.

2:10–2:40 pm

Coffee and discussion:

“The Future of IDIR”, Part 1 – group discussion.

2:40 pm

Discussion:

“The Future of IDIR”, Part 2 – reporting.

2:50 pm

Dr ANNE SMITH, School of Biology, University of St Andrews
(vas1@st-andrews.ac.uk)

“Data science in Bayesian networks for biological systems”.

3:10 pm

Dr SOPHIE MULLINS, Knowledge Transfer Centre, University of St Andrews
(sophie.mullins@st-andrews.ac.uk)

“Data-directed historical research. The Parisian printing industry of the early C16th”.

3:30 pm

Dr TANJA VAN MOURIK, School of Chemistry, University of St Andrews
(tanja.vanmourik@st-andrews.ac.uk)

“Veracity through variety (of methods): simulating dipeptides with little volume”.

3:50 pm

Dr UTA HINRICHS, School of Computer Science (uh3@st-andrews.ac.uk),
and Dr KONRAD LAWSON, School of History (kml8@st-andrews.ac.uk),
University of St Andrews

“Announcement: Future discussion series”.

3:55–4:00 pm

Comfort break.

4:00 pm

Discussion panel.

Members: all speakers, above, plus –

Professor AARON QUIGLEY, School of Computer Science
(aquigley@st-andrews.ac.uk), and Professor ANDREW CAMERON FRSE,
School of Physics and Astronomy, University of St Andrews
(acc4@st-andrews.ac.uk)

Chair: Professor DEREK WOOLLINS, Vice Principal (Research),
University of St Andrews (vpresearch@st-andrews.ac.uk).

The panel will discuss the following questions:

- (1) “How has Data Science changed your field?”
- (2) “What would you like Data Science to do for your field in the future?”

In addition: *all members of the audience are encouraged to ask questions, suggest topics on Data Science and join in the discussion.*

4:55 pm

Wrap up.

5:00–6:00 pm

Drinks reception.

The Summer of V's

Organising Committee: Dr Daniel Barker, chair (db60@st-andrews.ac.uk), Professor Aaron Quigley (aquigley@st-andrews.ac.uk) and Professor Simon Dobson (Simon.Dobson@st-andrews.ac.uk).

Summer of V's: <http://www.idir.st-andrews.ac.uk/vs>

IDIR – Institute for Data-Intensive Research: <http://www.idir.st-andrews.ac.uk>

Acknowledgements

This symposium was generously funded by the University of St Andrews and the St Andrews University Bioinformatics Unit.

University of St Andrews: <http://www.st-andrews.ac.uk>

Bioinformatics Unit: <http://bioinformatics.st-andrews.ac.uk>

Freedom and Generalization (St Andrews, Summer of V's, Volume)

Dr Amit Pundik, Tel Aviv University, amitp@post.tau.ac.il

- Introducing the problem
 - Many generalizations are unobjectionable: a dinner invitation, Twelve Angry Men, average longevity
 - Any inference from a known to an unknown fact seems to require a generalization.
 - However, some generalizations are intuitively objectionable: racist; crime-rates.
 - Such generalizations are not merely insufficient: they also seem *inadmissible*.
 - Interestingly, the appropriateness of a generalization depends heavily on the context in which it is used: cf crime-rates in criminal conviction, policing and medical treatment.
- Existing accounts are unable to distinguish objectionable and unobjectionable generalizations.
 - Epistemic accounts: first and second order, objectionable generalizations lack weight, appropriate causal connection, case-specificity, inability to provide the best explanation, vulnerability to the problem of the reference class, and sensitivity to the truth.
 - Contextualist accounts: misleading use of numbers, legitimacy of the legal system, the over-transparency of standards of proof, equality between litigants, and the individuality and autonomy of the litigant.
- The culpability account: seeks to connect the problem of generalization to the issue of *free will*.
 - Some generalizations are objectionable because:
 - They require presupposing that the individual's behaviour is determined by a certain *causal* factor.
 - According to libertarian and some compatibilist accounts of free will, being exposed to that type of causal factor renders the individual's behaviour unfree.

- Yet, in some contexts, such as when *attributing culpability*, it is necessary to presuppose that the individual is free to determine her own behaviour.
 - Using these generalizations in such a context thus involves *inconsistent* presuppositions.
- Why generalizations about human behaviour have to be causal?
 - The claim: an inference is valid only if it presupposes the *existence* of a causal generalization (even if the causal relation is indirect or the cause itself is left unspecified).
 - The general argument - *spurious correlations*:
 - Spurious correlations do not reflect an actual connection (causal or not) between the types of fact.
 - They seem to be everywhere: e.g. the seemingly-perfect correlation between divorce rate in Maine and per capita consumption of margarine in the US.
 - Since each case consists of numerous details (most are unimportant, of course), one could go through a vast number of facts until a correlation is found.
 - E.g. one might find a correlation between certain human behaviour and the second (or third) letter of the person's great aunt's surname.
 - If only causal generalizations are valid for inferential purposes, one could apply one's methods to distinguish between causal and non-causal connection (e.g. Markov Condition, Bayesian Nets).
 - However, if one denies that a valid inference requires a causal generalization, one ought to find how to distinguish between informative and spurious correlations.
 - Note that mere statistical significance will not do: trying a sufficiently-large number of variables using sufficiently-large databases would eventually generate statistically-significant yet spurious generalizations.
 - One might hope that such absurd statistically-significant correlations simply do not exist.

- But here one hopes that genuine correlations must “make sense”, and what would such an explanation be if not causal or causal-like? (David Lewis)
 - The specific argument – *inside vs outside the jar*
 - Denying an underlying causal connection is easier when the generalization is extracted from a group of cases to which the specific case at hand *belongs*:
 - For example: the gatecrasher paradox, blue balls in a jar.
 - By contrast, generalizations used in legal fact-finding are almost always extracted from a group of cases which does not include the specific case.
 - For example: average longevity, Twelve Angry Men, crime-rates.
 - If the case at hand is included in the generalization, it might be possible to use it without presupposing anything about the relation between the *types* of fact.
 - E.g. it might be possible to infer that the probability of a ball’s being blue is 50.1% without presupposing any causal connection between “being in that jar” and “being blue”.
 - Yet inferring the probability of a ball’s being blue from the proportion of blue balls in *another* jar requires presupposing a substantial relation between “being in a jar” and “being blue”.
 - Similarly, inferring that someone is likelier to commit a violent crime from past crime-rates in his neighbourhood requires presupposing a substantial relation between “residing in that neighbourhood” and “acting violently”.
 - And again, if this substantial relation is not causal or causal-like, what else could it be?
- The rest of the argument:
 - For simplicity, assume a libertarian approach to free will: for libertarians, the very presupposition of a causal factor outside the agent’s control suffices to render the behaviour unfree.

- This remains true even if the causal factor is unspecified or unknown.
 - How the argument works:
 - Deterministic generalizations: the skin mark example.
 - Indeterministic generalizations:
 - Under a subjectivist interpretation of probability
 - Under an objectivist interpretation of probability
- The scope of the culpability account: three additive conditions
 - The inferred (unknown) fact is *human behaviour*;
 - Human rather than an inanimate object.
 - Example: an explosion in a factory: a bomb or a malfunction?
 - Human's behaviour rather than other human properties
 - Example: longevity
 - The inferred behaviour is *culpable*;
 - Behaviour ought to be *free*: Donna and the rape-drug
 - behaviour ought to be *wrongful*: Donna and coffee-drinking
 - The culpable behaviour was *caused* by a property the person shares with other people to whom the generalization applies.
 - Unobjectionable if the culpable behaviour caused, rather than was caused by, the shared property.
 - Example: running away from the crime scene
- Test cases:
 - Unobjectionable generalizations existing accounts are unable to explain:
 - Assessing the credibility of an eyewitness
 - DNA Evidence
 - Objectionable generalizations existing accounts have tried to explain
 - Gatecrasher and blue/red bus paradoxes
- Extending the argument
 - Generalizations used to decide whom to stop and search (racial profiling)
 - Generalizations used to rule out non-culpable causes (Sally Clark)
 - Generalizations used to infer consent to sexual relations (previous sexual history)