

Knowledge Discovery by Design

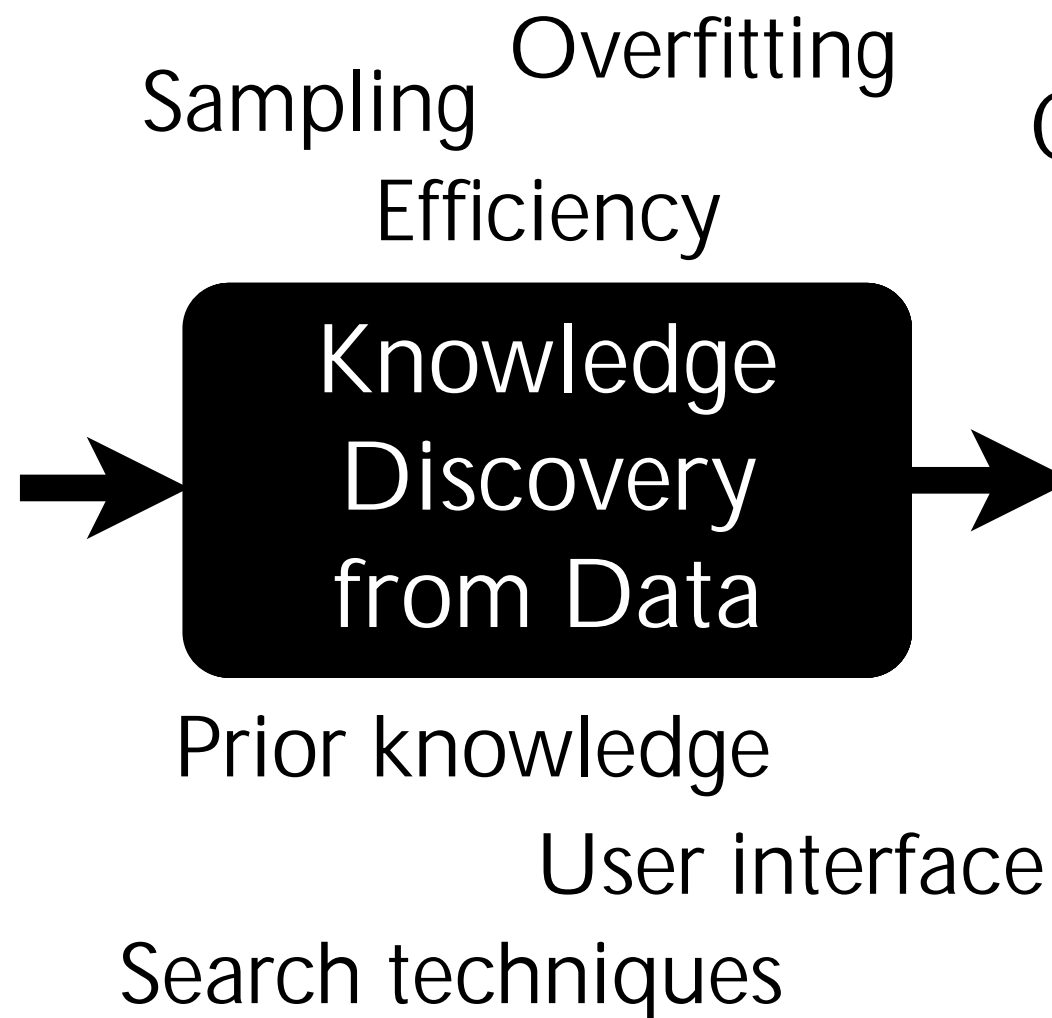
David Jensen

Knowledge Discovery Laboratory
Department of Computer Science
University of Massachusetts Amherst



Numeric Data
Symbolic Data
Relational DB
Text
Images
Temporal Data
Spatial Data
.....

These things
are
data...



Classification Trees
Association Rules
Neural Nets
Bayes Nets
Linear Equations
Horn Clauses
...

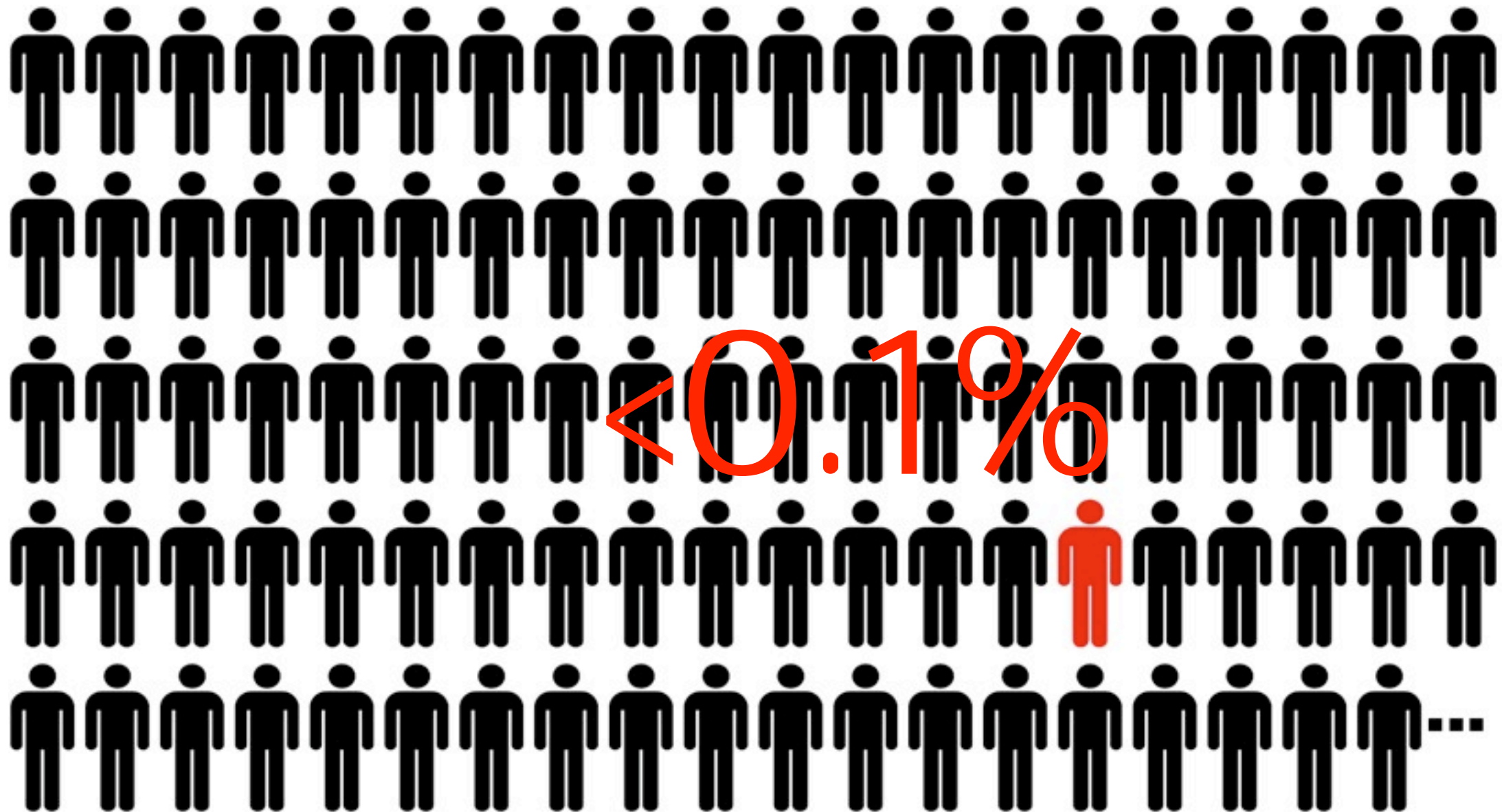
...but are
these things
knowledge?

Example: Stock Fraud Detection

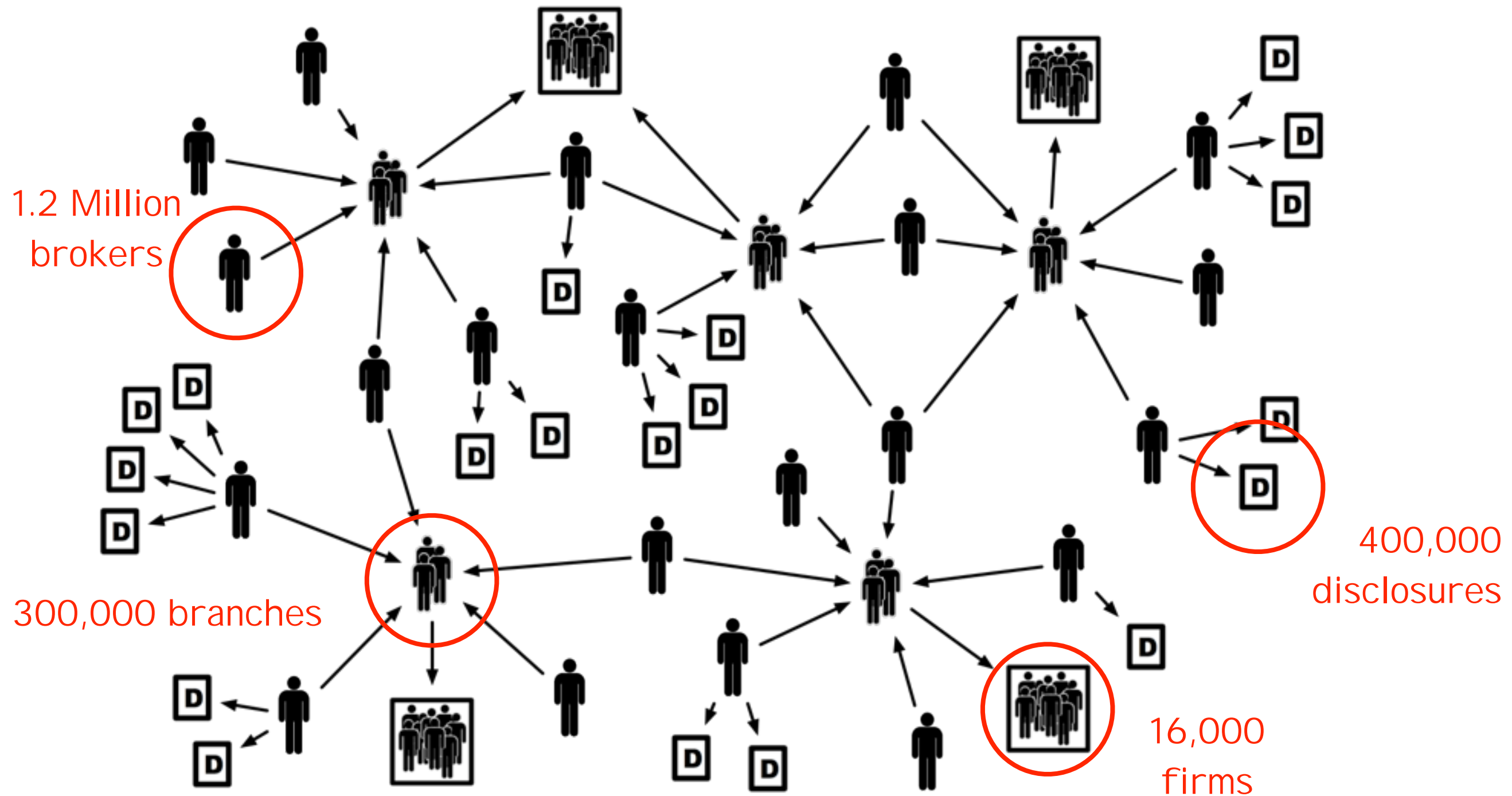
The Financial Industry Regulatory Authority (FINRA) regulates virtually every securities firm doing business with the US public. They register member firms, write rules to govern their behavior, examine them for compliance and discipline those that fail to comply.



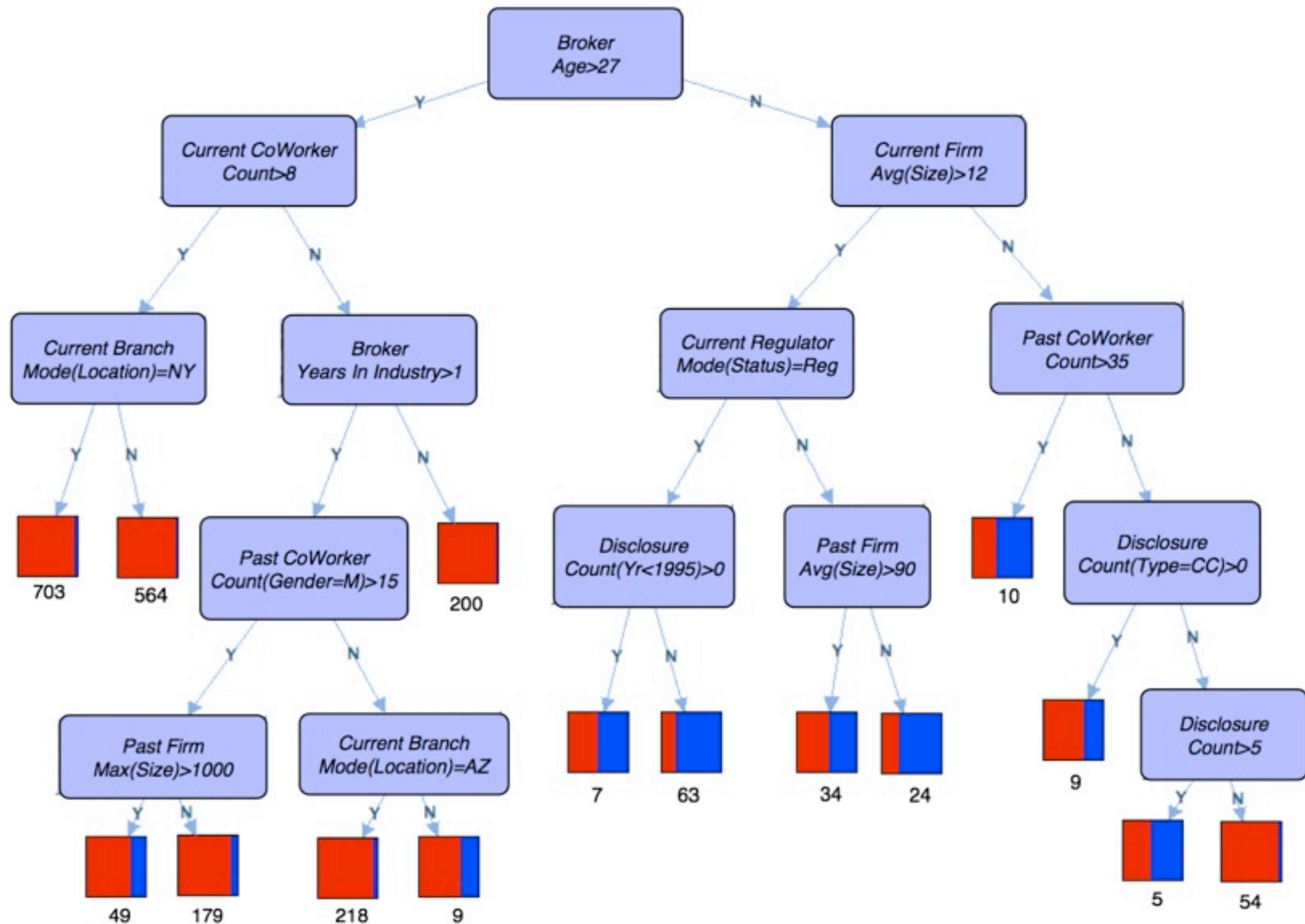
Broker fraud is rare

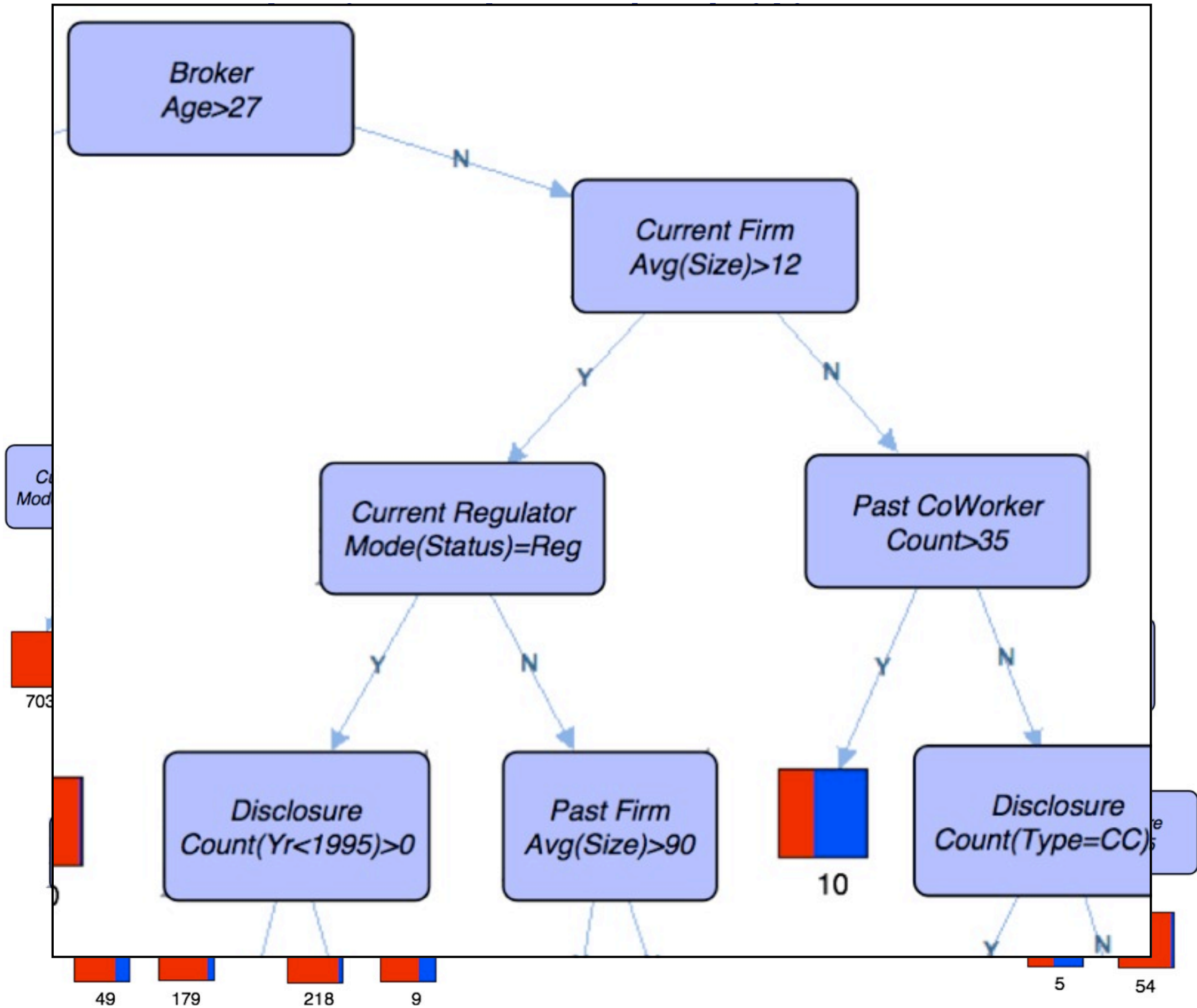


FINRA data

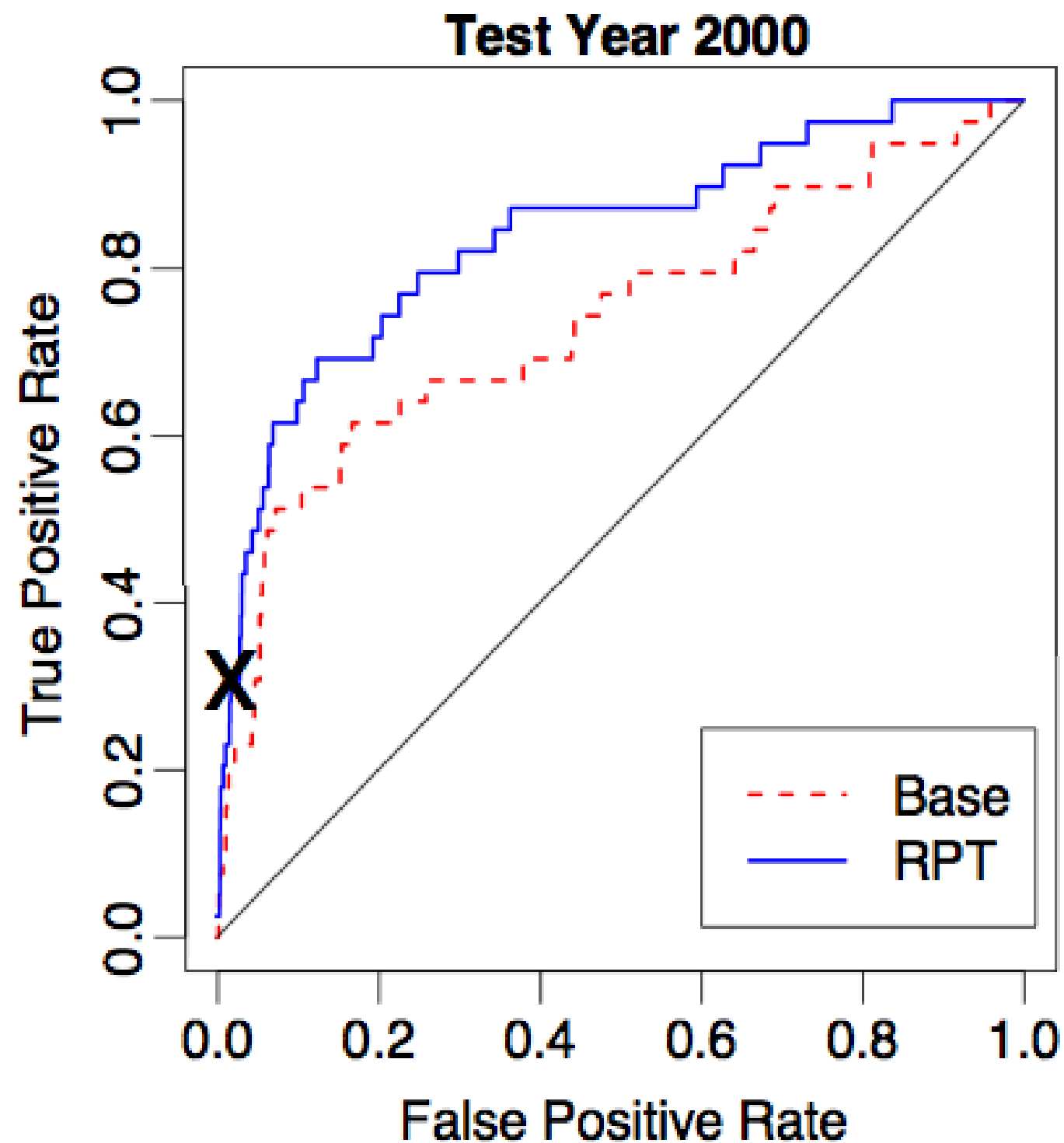


Relational probability trees

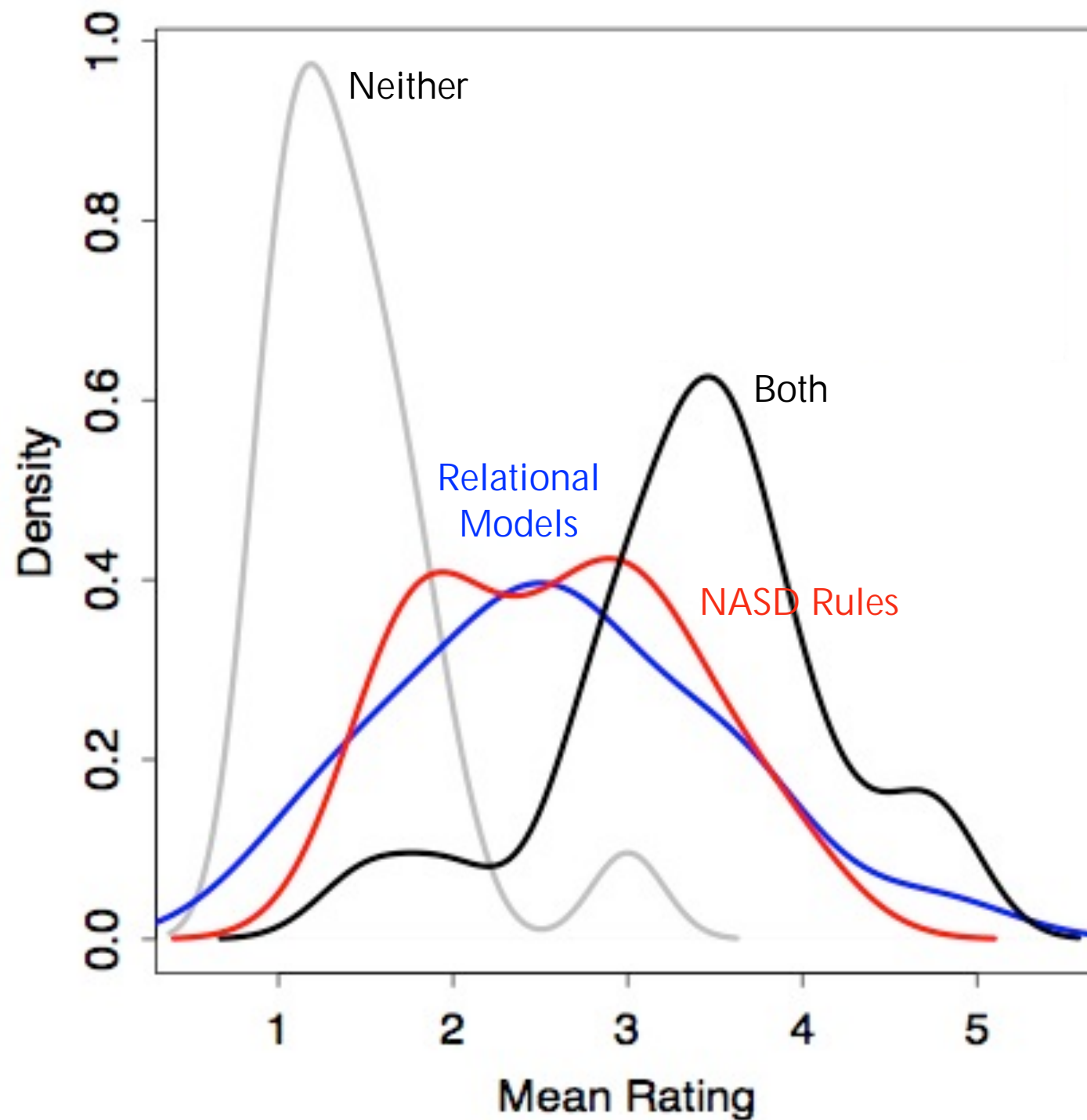




Results — Objective accuracy

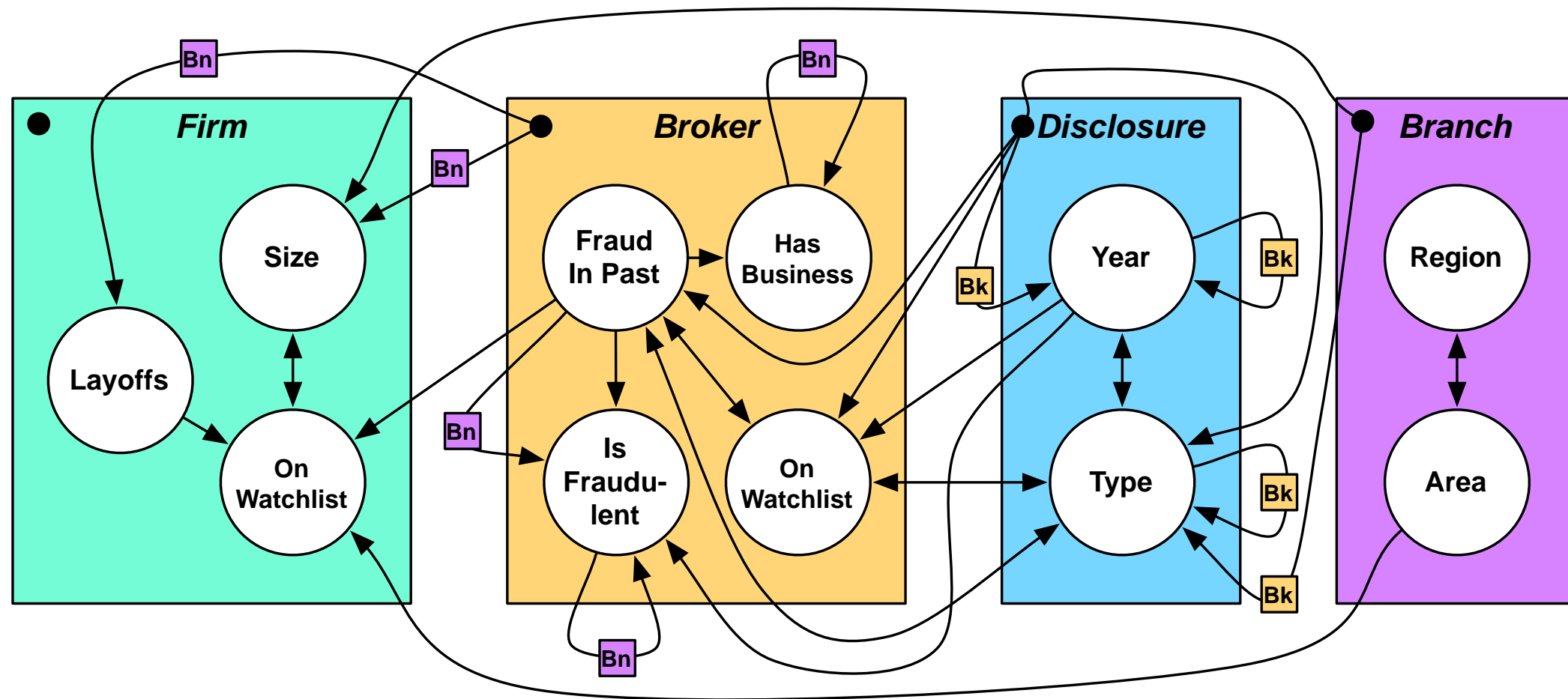


Results — Aiding examiners



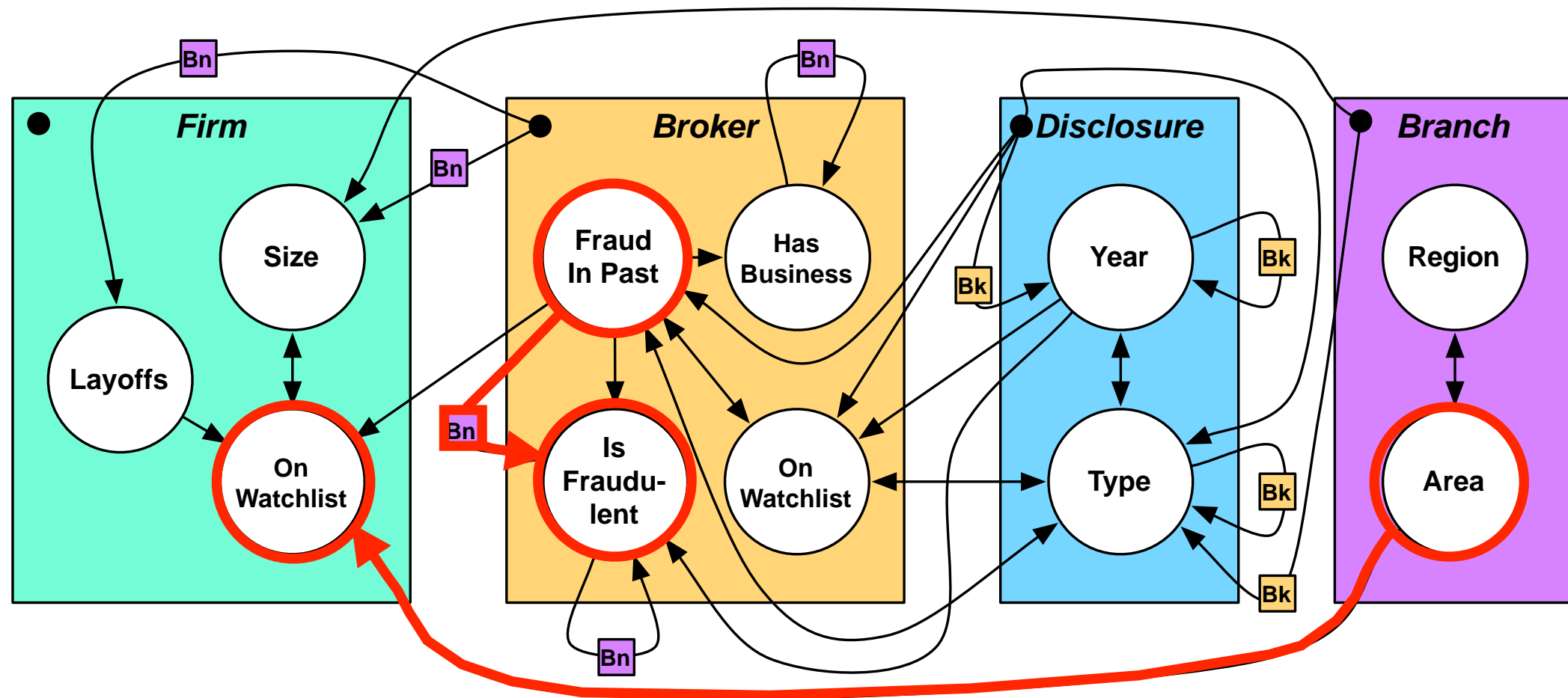
(Neville et al. 2005)

Relational dependency networks



(Neville & Jensen 2004, 2006)

Experts interpret models causally



It's what they do

It's what they want to do

It's what they should want to do

Assertion 1

More research should focus on
representing, learning, and managing
causal knowledge

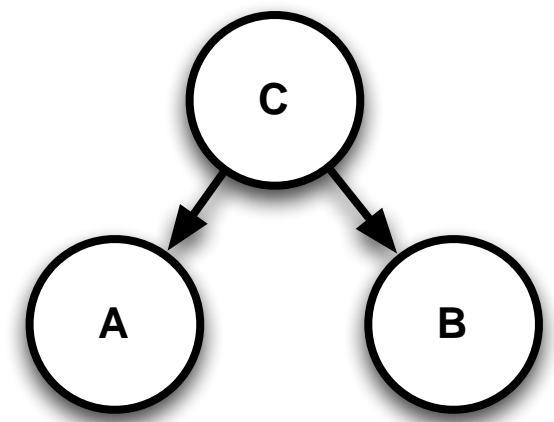
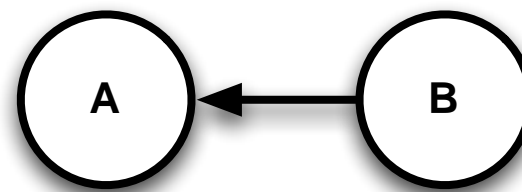
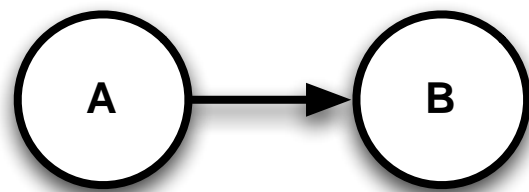
What is causality?

“The paradigmatic assertion in causal relationships is that manipulation of a cause will result in the manipulation of an effect. ... Causation implies that by varying one factor, I can make another vary.”

– Cook & Campbell (1979)

Why discover causality?

- A frequent goal of knowledge discovery projects is actionable knowledge.
- Statistical association between A and B is insufficient to distinguish among different causal models.



- Each causal model implies different actions, if we wish to influence the value of B.

How is causal knowledge different?

Domain of Analysis

Economics
and finance

Consumer response

Elections

Social science

Citation analysis

Biological and
ecological modeling

Domain of Action

Fiscal and
regulatory policy

Marketing

Campaign strategy

Social policy

Science policy

Prevention, treatment,
and remediation

Assertion 2

Learning causal knowledge from data
poses significant challenges

Adolescent Sexual Debut and Later Delinquency

Stacy Armour · Dana L. Haynie

Received: 31 May 2006 / Accepted: 7 August 2006 / Published online: 13 September 2006
© Springer Science+Business Media, Inc. 2006

Abstract Does sexual debut (i.e., experiencing sexual intercourse for the first time) increase the risks of participating in later delinquent behavior? Does this risk increase if adolescents experience early sexual debut relative to the timing experienced by one's peers? Although many factors have been linked to sexual debut, little research has examined whether sexual initiation is linked to later delinquency. Using data on adolescents participating in the National Longitudinal Study of Adolescent Health (2007), we examine the interconnectedness of sexual debut and later delinquency. In addition, we examine the role of timing of sexual debut on later delinquency. Sexual debut is associated with delinquency one year later. In addition, those adolescents who experience early sexual debut relative to their peers are at higher risk of experiencing delinquency compared to those who experience late sexual debut. Moreover, the protective effect of late sexual debut appears to persist for several years. Findings are interpreted by drawing on developmental theory and life course research.

Introduction

Interest in the timing, prediction, and consequences of sexual initiation continues to garner much research attention. In part, this interest has resulted from the steadily declining age at which adolescents begin having sex and the finding that in the U.S. most adolescents report having had

sex by the time they graduate from high school (Singh and Darroch, 2000). Indeed, recent political discussion has focused on teen sexual activity and the promotion of abstinence education because of the concern that sex outside of marriage is likely to have deleterious consequences for youth. However, despite the common assumption that sexual initiation in general, and sexual debut in particular, is likely to result in poor outcomes, little research to our knowledge has examined this claim. Although many studies have examined the relationship between sexual debut and later delinquency, little research has focused on the timing of sexual initiation. Sexual initiation is linked to problem behavior and involvement in delinquency later in the life course. The purpose of this research is to examine the association between sexual debut and later adolescent delinquency, with a particular focus on early sexual debut.

Although engaging in delinquency is not uncommon during adolescence (Moffitt, 1993) there is evidence that for some adolescents, delinquency can have lasting effects reducing social and human capital that can then result in diminished economic and social opportunities in young adulthood (Hagan, 1998; Hagan et al., 1996). This may be particularly the case if adolescents engage in delinquency at a young age. For instance, if early sexual debut initiates or results in elevated delinquency. Therefore, examining the associations between sexual debut and later delinquency is warranted.

Developmental theory is instrumental in explaining how early sexual initiation can open the doorway to other behaviors. Of primary interest here, developmental theory and life course research suggest that timing of sexual debut is

¹ By sexual debut we refer to transitioning from virgin to non-virgin status based on self-reported information on the experience of sexual intercourse.

(Armour & Haynie 2007)

Rethinking Timing of First Sex and Delinquency

K. Paige Harden · Jane Mendle · Jennifer E. Hill ·
Eric Turkheimer · Robert E. Emery

Received: 24 July 2007 / Accepted: 11 September 2007
© Springer Science+Business Media, LLC 2007

Abstract The relation between timing of first sex and later delinquency was examined using a genetically informed sample of 534 same-sex twin pairs from the National Longitudinal Study of Adolescent Health, who were assessed at three time points over a 7-year interval. Genetic and environmental differences between families were found to account for the association between earlier age at first sex and increases in delinquency. Controlling for these genetic and environmental effects, individuals using a quasi-experimental design, predicted lower levels of delinquency later in life. The current study is contrasted with previous research with non-genetically informative samples. **Keywords** Adolescent sexual behavior · First sex · Sexual debut · Delinquency · Behavior genetics

Introduction

Armour and Haynie recently published a report on "Adolescent sexual debut and later delinquency" in the February 2007 issue of *Journal of Youth and Adolescence*. The authors' stated intent was to examine whether earlier "sexual debut," relative to peers, increases the

risk for delinquent behavior. The authors successfully utilized a large and complex data set—the National Longitudinal Study of Adolescent Health—to demonstrate that adolescents who have experienced first sex in the past year exhibited higher levels of delinquency compared to adolescents who remained virgins. In addition, adolescents who experienced first sex earlier than their same-school peers exhibited an increase in delinquency later in life. Overall, this was an interesting and important finding among variables of considerable developmental significance. We have no qualms with either the data or the execution of the analysis itself. Nevertheless, Armour and Haynie's (2007) interpretation of their results typifies a problem commonly encountered in research on adolescent development—drawing unwarranted causal conclusions from non-experimental data.

Armour and Haynie are no doubt well aware of the logical pitfalls of concluding that *X* causes *Y* from non-experimental data. Yet, like many researchers, they interpret an epidemiological association between the timing of first sex and delinquency—with little or no qualification—as evidence that early first sex causes later delinquency. For example, they conclude that "the timing...of events such as sexual activity has profound consequences" (p. 149) and that "experiencing early or late sexual debut continues to have consequences for delinquent behavior occurring in young adulthood" (p. 150). An alternative explanation for their findings, of course, is that a third variable, either environmental or genetic, is associated with both timing of first sex and later delinquent behavior. That is, the association between early sex and later delinquency reported by Armour and Haynie (2007) could be an artifact of uncontrolled confounds.

K. P. Harden (✉) · J. Mendle · J. E. Hill · E. Turkheimer ·
R. E. Emery
Department of Psychology, University of Virginia,
Charlottesville, VA 22904-4408, USA
e-mail: paigharden@virginia.edu

(Harden et al. 2007)

Adolescent Sexual Debut and Later Delinquency

Stacy Armour · Dana L. Haynie

Received: 31 May 2006 / Accepted: 7 August 2006 / Published online: 13 September 2006
© Springer Science+Business Media, Inc. 2006

Abstract Does sexual debut (i.e., experiencing sexual intercourse for the first time) increase the risks of participating in later delinquent behavior? Does this risk increase if adolescents experience early sexual debut relative to the timing experienced by one's peers? Although many factors have been linked to sexual debut, little research has examined whether sexual initiation is linked to later delinquency. Using data on adolescents participating in the National Longitudinal Study of Adolescent Health (1994–2007), we examine the interrelated associations between sexual debut and later delinquency. In addition, we examine the association to the role of timing of sexual debut. Experiencing sexual debut is associated with delinquency one year later. In addition, those adolescents who experience early sexual debut relative to their peers are at higher risk of experiencing delinquency compared to those who experience late sexual debut. Findings are interpreted by drawing on developmental theory and life course research.

Introduction

Interest in the timing, prediction, and consequences of sexual initiation continues to garner much research attention. In part, this interest has resulted from the steadily declining age at which adolescents begin having sex and the finding that in the U.S. most adolescents report having had

sex by the time they graduate from high school (Singh and Darroch, 2000). Indeed, recent political discussion has focused on teen sexual activity and the promotion of abstinence education because of the concern that sex outside of marriage is likely to have deleterious consequences for youth. However, despite the common assumption that sexual initiation in general, and sexual debut in particular, is likely to result in poor outcomes, little research to our knowledge has examined this claim. Although many studies have examined the association between sexual debut and later delinquency, little research has focused on the timing of sexual initiation. Sexual initiation is linked to problem behavior and involvement in delinquency later in the life course. The purpose of this research is to examine the association between sexual debut and later adolescent delinquency, with a particular focus on early sexual debut.

Although engaging in delinquency is not uncommon during adolescence (Moffitt, 1993) there is evidence that for some adolescents, delinquency can have lasting effects reducing social and human capital that can then result in diminished economic and social opportunities in young adulthood (Hagan, 1998; Hagan et al., 1996). This may be particularly the case if adolescents engage in delinquency at a young age. For instance, if early sexual debut initiates or results in elevated delinquency. Therefore, examining the associations between sexual debut and later delinquency is warranted.

Developmental theory is instrumental in explaining how early sexual initiation can open the doorway to other behaviors. Of primary interest here, developmental theory and life course research suggest that timing of sexual debut is

¹ By sexual debut we refer to transitioning from virgin to non-virgin status based on self-reported information on the experience of sexual intercourse.

(Armour & Haynie 2007)

Rethinking Timing of First Sex and Delinquency

K. Paige Harden · Jane Mendle · Jennifer E. Hill ·
Eric Turkheimer · Robert E. Emery

Received: 24 July 2007 / Accepted: 11 September 2007
© Springer Science+Business Media, LLC 2007

Genetics and Family Environment

Abstract The relation between timing of first sex and later delinquent behavior was examined using a genetically informed sample of 534 same-sex twin pairs from the National Longitudinal Study of Adolescent Health, who were assessed at three time points over a 7-year interval. Genetic and environmental differences between families were found to account for the association between earlier age at first sex and increases in delinquency. After controlling for these genetic and environmental confounds using a quasi-experimental design, earlier age at first sex predicted lower levels of delinquency in adolescence. The current study is contrasted with previous research with non-genetically informative samples (Armour and Haynie (2007, *Journal of Youth and Adolescence* 36:141–152). Results suggest a more nuanced interpretation of the meaning and consequences of adolescent sexual debut, which is commonly put forth in the literature.

Keywords Adolescent sexual behavior · First sex · Sexual debut · Delinquency · Behavior

Introduction

Armour and Haynie recently published a report on “Adolescent sexual debut and later delinquency” in the February 2007 issue of *Journal of Youth and Adolescence*. The authors’ stated intent was to examine whether earlier “sexual debut,” relative to peers, increases the

risks for delinquent behavior. The authors successfully analyzed a large and complex data set—the National Longitudinal Study of Adolescent Health—to demonstrate that adolescents who have experienced first sex in early adolescence exhibited higher levels of delinquency than adolescents who remained virgins. In addition, adolescents who experienced first sex earlier than their school peers exhibited an increase in delinquency later in adolescence and early adulthood, controlling for various delinquency. Overall, this was an interesting analysis of associations among variables of considerable developmental importance. We have no qualms with either the execution of the analysis itself, or the interpretation of the results. A problem commonly encountered in research on adolescent development—drawing causal conclusions from non-experimental data—was not present.

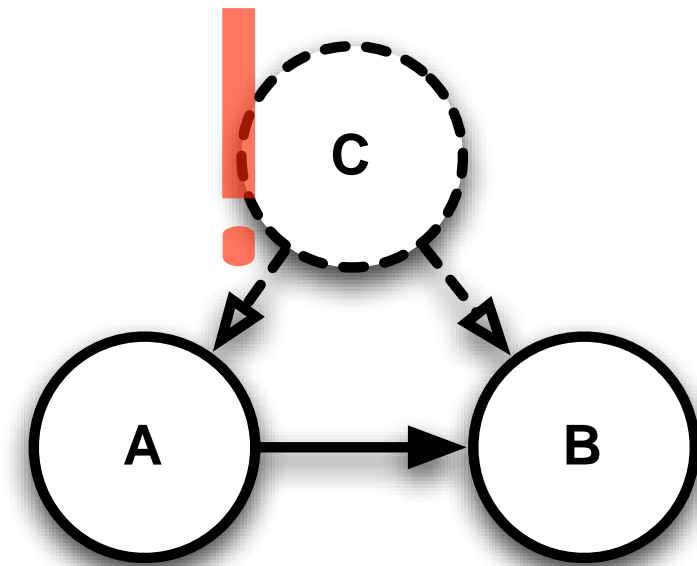
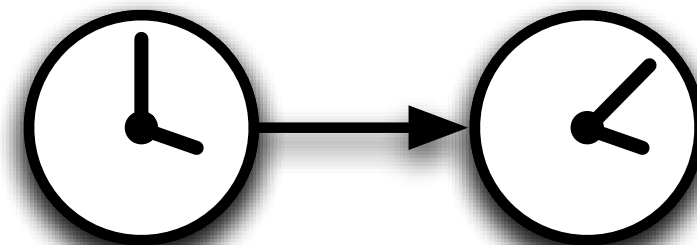
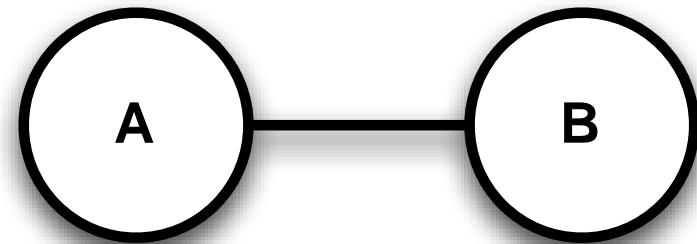
Armour and Haynie are no doubt well aware of the logical pitfalls of inferring causality from non-experimental data. Yet, like many researchers, they interpret an epidemiological association between the timing of first sex and delinquency—with little or no qualification—as evidence that early first sex causes later delinquency. For example, they conclude that “the timing...of events such as sexual activity has profound consequences” (p. 149) and that “experiencing early or late sexual debut continues to have consequences for delinquent behavior occurring in young adulthood” (p. 150). An alternative explanation for their findings, of course, is that a third variable, either environmental or genetic, is associated with both timing of first sex and later delinquent behavior. That is, the association between early sex and later delinquency reported by Armour and Haynie (2007) could be an artifact of uncontrolled confounds.

K. P. Harden (✉) · J. Mendle · J. E. Hill · E. Turkheimer · R. E. Emery
Department of Psychology, University of Virginia,
Charlottesville, VA 22904-4408, USA
e-mail: kpharden@virginia.edu

(Harden et al. 2007)

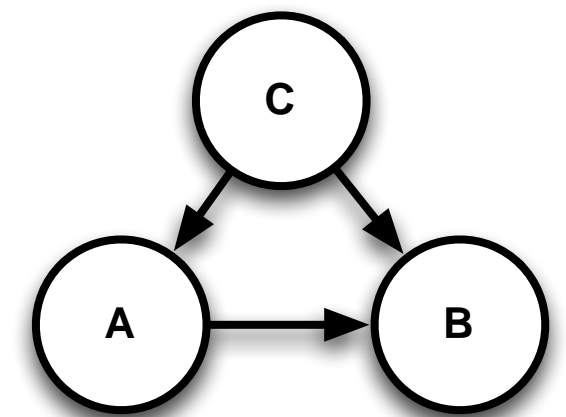
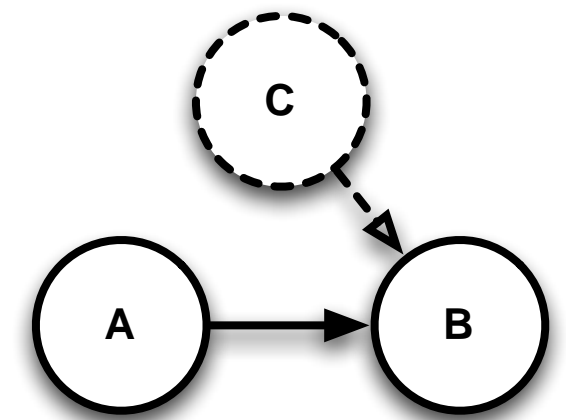
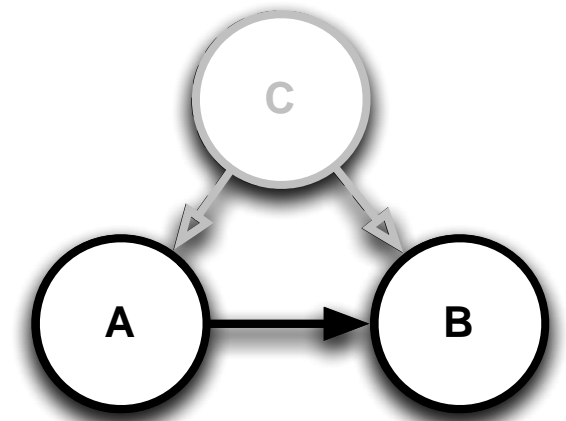
Conditions for causal inference

- Association
- Direction
- Elimination of potential common causes



Eliminating common causes

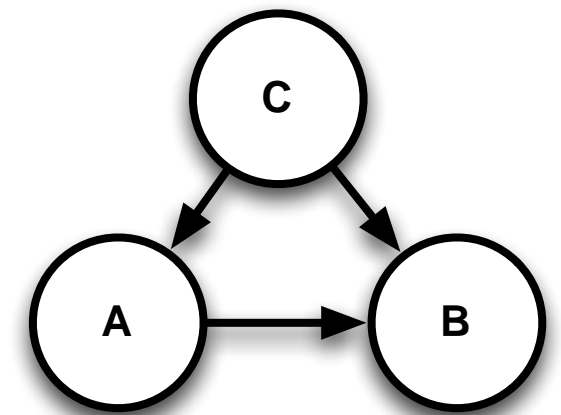
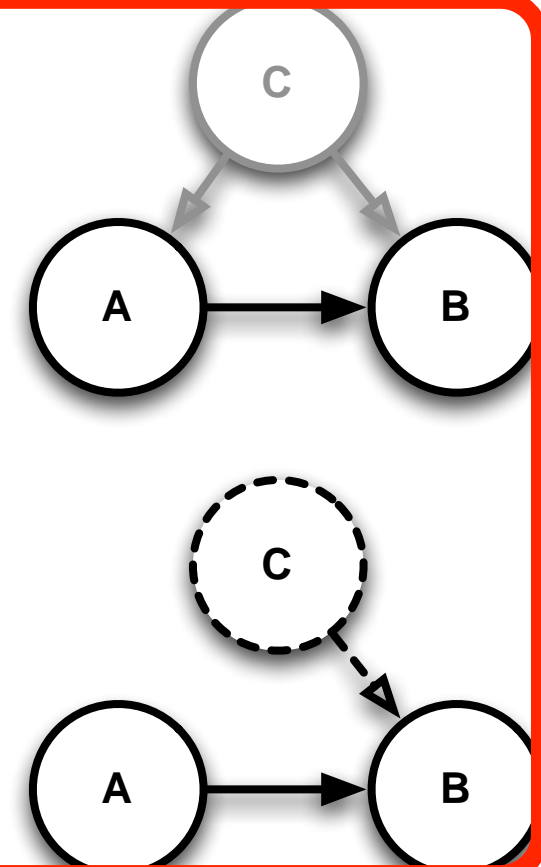
- Control — Hold potential common causes constant so they cannot affect outcomes (Bacon 1620)
- Randomization — Assign treatments randomly so potential common causes cannot systematically affect outcomes (Fisher 1925)
- Modeling — Measure, model and mathematically remove effects of potential common causes (Rubin 1974; Spirtes, Glymour & Scheines 1993; Pearl 2000)



Eliminating common causes

Only useful for experiments

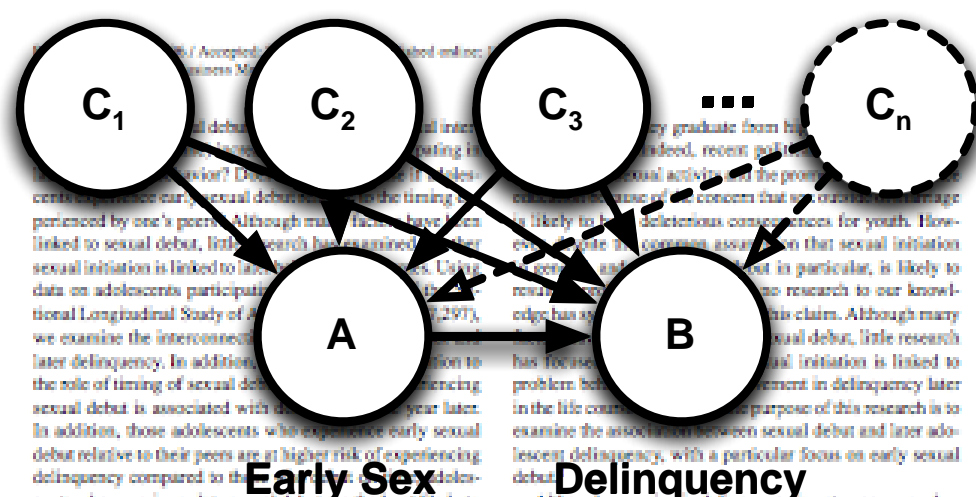
- Control — Hold potential common causes constant so they cannot affect outcomes (Bacon 1620)
- Randomization — Assign treatments randomly so potential common causes cannot systematically affect outcomes (Fisher 1925)
- Modeling — Model and mathematically remove effects of potential common causes (Rubin 1974; Spirtes, Glymour & Scheines 1993; Pearl 2000)



Adolescent Sexual Debut and Later Delinquency

Stacy Armour · Dana L. Haynie

Modeling



Introduction

Interest in the timing, prediction, and consequences of sexual initiation continues to garner much research attention. In part, this interest has resulted from the steadily declining age at which adolescents begin having sex and the finding that in the U.S. most adolescents report having had

sexual intercourse by age 18 (Hagan et al., 1998). Although engaging in delinquency is not uncommon during adolescence (Moffitt, 1993) there is evidence that for some adolescents, delinquency can have lasting effects reducing social and human capital that can then result in diminished economic and social opportunities in young adulthood (Hagan, 1998; Hagan et al., 1998). This may be particularly the case if adolescents engage in delinquency at a young age. For instance, if early sexual debut initiates or results in elevated delinquency. Therefore, examining the associations between sexual debut and later delinquency is warranted.

Developmental theory is instrumental in explaining how early sexual initiation can open the doorway to other behaviors. Of primary interest here, developmental theory and life course research suggest that timing of sexual debut is

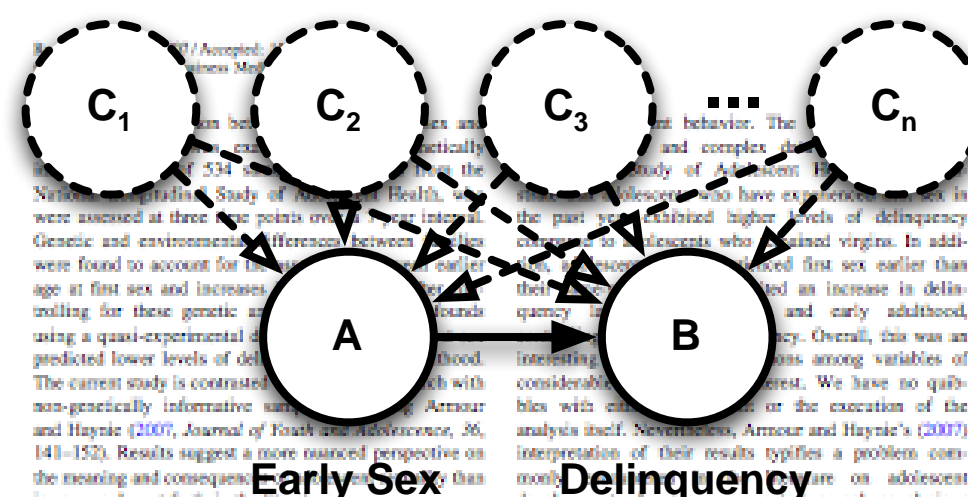
¹ By sexual debut we refer to transitioning from virgin to non-virgin status based on self-reported information on the experience of sexual intercourse.

S. Armour (✉) · D. L. Haynie
Department of Sociology, The Ohio State University,
360 Bricker Hall, 180 N. Oval Mall, Columbus, OH 43210, USA
e-mail: armour.12@osu.edu

(Armour & Haynie 2007)

Rethinking Timing of First Sex and Delinquency

K. Paige Harden · Jane Mendle · Jennifer E. Hill ·
Eric Turkheimer · Robert E. Emery



Keywords Adolescent sexual behavior · First sex · Sexual debut · Delinquency · Behavior genetics

Introduction

Armour and Haynie recently published a report on "Adolescent sexual debut and later delinquency" in the February 2007 issue of *Journal of Youth and Adolescence*. The authors' stated intent was to examine whether earlier "sexual debut," relative to peers, increases the

Armour and Haynie are no doubt well aware of the logical pitfalls of concluding that *X* causes *Y* from non-experimental data. Yet, like many researchers, they interpret an epidemiological association between the timing of first sex and delinquency—with little or no qualification—as evidence that early first sex causes later delinquency. For example, they conclude that "the timing...of events such as sexual activity has profound consequences" (p. 149) and that "experiencing early or late sexual debut continues to have consequences for delinquent behavior occurring in young adulthood" (p. 150). An alternative explanation for their findings, of course, is that a third variable, either environmental or genetic, is associated with both timing of first sex and later delinquent behavior. That is, the association between early sex and later delinquency reported by Armour and Haynie (2007) could be an artifact of uncontrolled confounds.

K. P. Harden (✉) · J. Mendle · J. E. Hill · E. Turkheimer ·
R. E. Emery
Department of Psychology, University of Virginia,
Charlottesville, VA 22904-4408, USA
e-mail: paigeharden@virginia.edu

(Harden et al. 2007)

Statistical control

- Armour and Haynie modeled a large number of potential influences on adolescent behavior.

Female	-0.065***
Receipt of public assistance	0.031
Delinquency at W I	0.151***
Parental support	-0.063
Dating	-0.024
Importance of religion	-0.042
Virginity pledge	-0.206*
Relative pubertal status	-0.037

- Even after accounting for these potential causes, a statistical association remained between early sexual activity and delinquency.

Modeling can be difficult

- Assumption of causal sufficiency —
All potential common causes are represented in the data and measured accurately
- Assumption of correctly specified models —
The joint effects of all potential common causes are modeled accurately
- and there are additional issues...

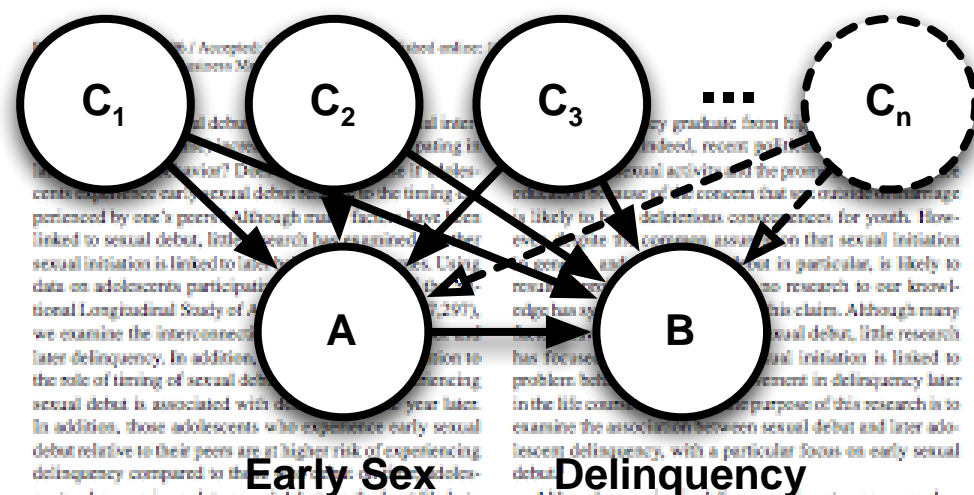
Modeling can be difficult

- Unresolvable causal dependencies — Large equivalence classes of models containing instances with different causal implications
- High sample complexity — Large data sets are needed to resolve some dependencies
- High computational complexity — Even if sufficiently large samples exist, effective structure learning can be intractable for reasonably sized networks

Adolescent Sexual Debut and Later Delinquency

Stacy Armour · Dana L. Haynie

Modeling



Introduction

Interest in the timing, prediction, and consequences of sexual initiation continues to garner much research attention. In part, this interest has resulted from the steadily declining age at which adolescents begin having sex and the finding that in the U.S. most adolescents report having had

sexual debut by age 18 (Hagan et al., 1998). Although engaging in delinquency is not uncommon during adolescence (Moffitt, 1993) there is evidence that for some adolescents, delinquency can have lasting effects reducing social and human capital that can then result in diminished economic and social opportunities in young adulthood (Hagan, 1998; Hagan et al., 1998). This may be particularly the case if adolescents engage in delinquency at a young age. For instance, if early sexual debut initiates or results in elevated delinquency. Therefore, examining the associations between sexual debut and later delinquency is warranted.

Developmental theory is instrumental in explaining how early sexual initiation can open the doorway to other behaviors. Of primary interest here, developmental theory and life course research suggest that timing of sexual debut is

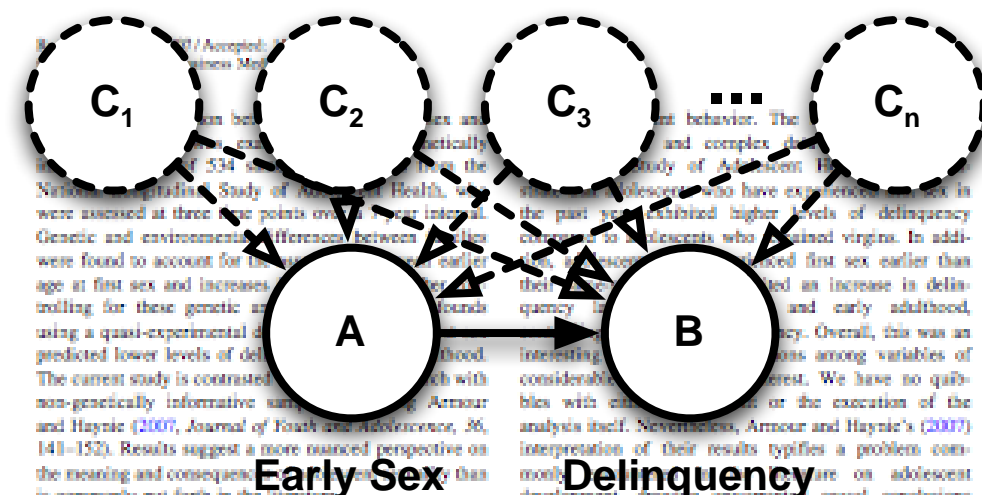
¹ By sexual debut we refer to transitioning from virgin to non-virgin status based on self-reported information on the experience of sexual intercourse.

S. Armour (✉) · D. L. Haynie
Department of Sociology, The Ohio State University,
360 Bricker Hall, 180 N. Oval Mall, Columbus, OH 43210, USA
e-mail: armour.12@osu.edu

(Armour & Haynie 2007)

Rethinking Timing of First Sex and Delinquency

K. Paige Harden · Jane Mendle · Jennifer E. Hill ·
Eric Turkheimer · Robert E. Emery



Keywords Adolescent sexual behavior · First sex · Sexual debut · Delinquency · Behavior genetics

Introduction

Armour and Haynie recently published a report on "Adolescent sexual debut and later delinquency" in the February 2007 issue of *Journal of Youth and Adolescence*. The authors' stated intent was to examine whether earlier "sexual debut," relative to peers, increases the

Armour and Haynie are no doubt well aware of the logical pitfalls of concluding that *X* causes *Y* from non-experimental data. Yet, like many researchers, they interpret an epidemiological association between the timing of first sex and delinquency—with little or no qualification—as evidence that early first sex causes later delinquency. For example, they conclude that "the timing...of events such as sexual activity has profound consequences" (p. 149) and that "experiencing early or late sexual debut continues to have consequences for delinquent behavior occurring in young adulthood" (p. 150). An alternative explanation for their findings, of course, is that a third variable, either environmental or genetic, is associated with both timing of first sex and later delinquent behavior. That is, the association between early sex and later delinquency reported by Armour and Haynie (2007) could be an artifact of uncontrolled confounds.

K. P. Harden (✉) · J. Mendle · J. E. Hill · E. Turkheimer ·
R. E. Emery
Department of Psychology, University of Virginia,
Charlottesville, VA 22904-4408, USA
e-mail: paigharden@virginia.edu

(Harden et al. 2007)

washingtonpost.com

Study Debunks Theory On Teen Sex, Delinquency

New Analyses Challenging Many Old Assumptions

By Rick Warren
Washington Post
Sunday, Nov. 11, 2007

Research
attention
lose their
to become
established
delinquency
curriculum

There was

Science News

[Share](#) [Blog](#) [Cite](#)

Early Teen Sex May Not Be A Path To Delinquency, Study Shows

ScienceDaily (Nov. 14, 2007) — A new study by University of Virginia clinical psychologists has found that teens who have sex at an early age may be less inclined to exhibit delinquent behavior in early adulthood than their peers who waited until they were older to have sex. The study also suggests that early sex may play a role in helping these teens develop better social relationships in early adulthood.

See also:

Health & Medicine

- [Teen Health](#)
- [Sexual Health](#)
- [Erectile Dysfunction](#)

Mind & Brain

- [Relationships](#)

The finding is published in the current online edition of the *Journal of Youth and Adolescence*, and runs counter to most assumptions that relate early teen sex to later drug use, criminality, antisocial behavior and emotional problems. The finding also contradicts parts of a study published earlier this year in the

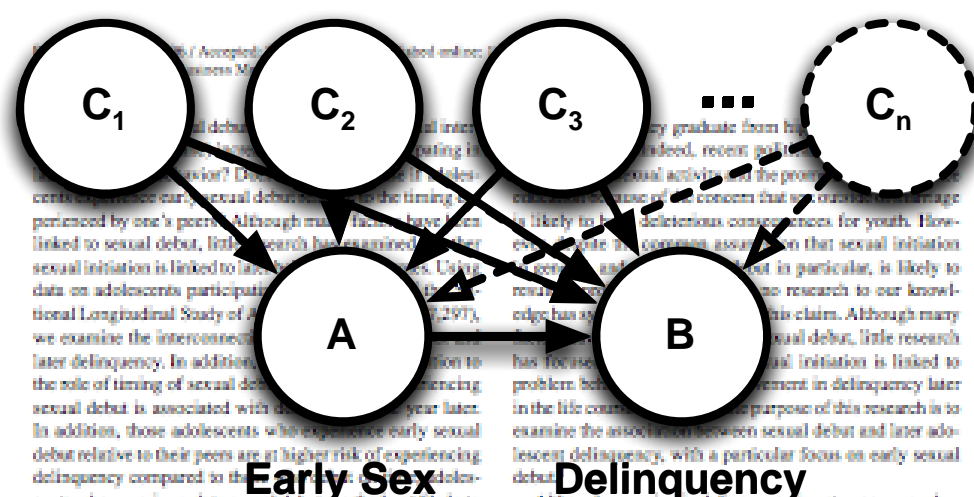


A new study by University of Virginia clinical psychologists has found that teens who have sex at an early age may be less inclined to exhibit delinquent behavior in early adulthood than their peers who waited until they were older to have sex. (Credit: iStockphoto/Pali Rao)

Adolescent Sexual Debut and Later Delinquency

Stacy Armour · Dana L. Haynie

Modeling



Introduction

Interest in the timing, prediction, and consequences of sexual initiation continues to garner much research attention. In part, this interest has resulted from the steadily declining age at which adolescents begin having sex and the finding that in the U.S. most adolescents report having had

sexual intercourse by age 18 (Moffitt, 1993). Although engaging in delinquency is not uncommon during adolescence (Moffitt, 1993) there is evidence that for some adolescents, delinquency can have lasting effects reducing social and human capital that can then result in diminished economic and social opportunities in young adulthood (Hagan, 1998; Hagan et al., 1996). This may be particularly the case if adolescents engage in delinquency at a young age. For instance, if early sexual debut initiates or results in elevated delinquency. Therefore, examining the associations between sexual debut and later delinquency is warranted.

Developmental theory is instrumental in explaining how early sexual initiation can open the doorway to other behaviors. Of primary interest here, developmental theory and life course research suggest that timing of sexual debut is

¹ By sexual debut we refer to transitioning from virgin to non-virgin status based on self-reported information on the experience of sexual intercourse.

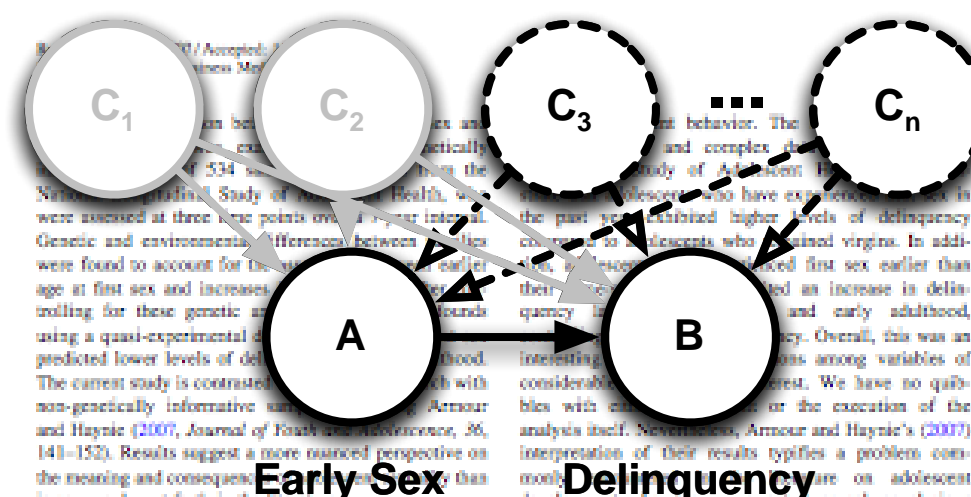
S. Armour (✉) · D. L. Haynie
Department of Sociology, The Ohio State University,
360 Bricker Hall, 180 N. Oval Mall, Columbus, OH 43210, USA
e-mail: armour.12@osu.edu

(Armour & Haynie 2007)

Rethinking Timing of First Sex and Delinquency

K. Paige Harden · Jane Mendle · Jennifer E. Hill ·
Eric Turkheimer · Robert E. Emery

Control



Keywords Adolescent sexual behavior · First sex · Sexual debut · Delinquency · Behavior genetics

Introduction

Armour and Haynie recently published a report on "Adolescent sexual debut and later delinquency" in the February 2007 issue of *Journal of Youth and Adolescence*. The authors' stated intent was to examine whether earlier "sexual debut," relative to peers, increases the

Armour and Haynie are no doubt well aware of the logical pitfalls of concluding that *X* causes *Y* from non-experimental data. Yet, like many researchers, they interpret an epidemiological association between the timing of first sex and delinquency—with little or no qualification—as evidence that early first sex causes later delinquency. For example, they conclude that "the timing...of events such as sexual activity has profound consequences" (p. 149) and that "experiencing early or late sexual debut continues to have consequences for delinquent behavior occurring in young adulthood" (p. 150). An alternative explanation for their findings, of course, is that a third variable, either environmental or genetic, is associated with both timing of first sex and later delinquent behavior. That is, the association between early sex and later delinquency reported by Armour and Haynie (2007) could be an artifact of uncontrolled confounds.

K. P. Harden (✉) · J. Mendle · J. E. Hill · E. Turkheimer ·
R. E. Emery
Department of Psychology, University of Virginia,
Charlottesville, VA 22904-4408, USA
e-mail: paigharden@virginia.edu

(Harden et al. 2007)

Assertion 3

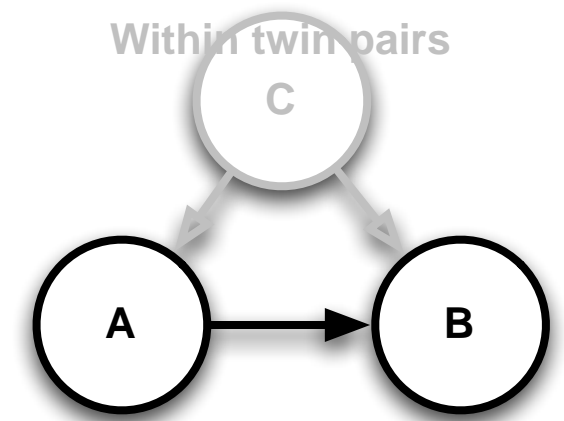
Quasi-experimental designs
provide a highly effective approach to
learning causal knowledge from data

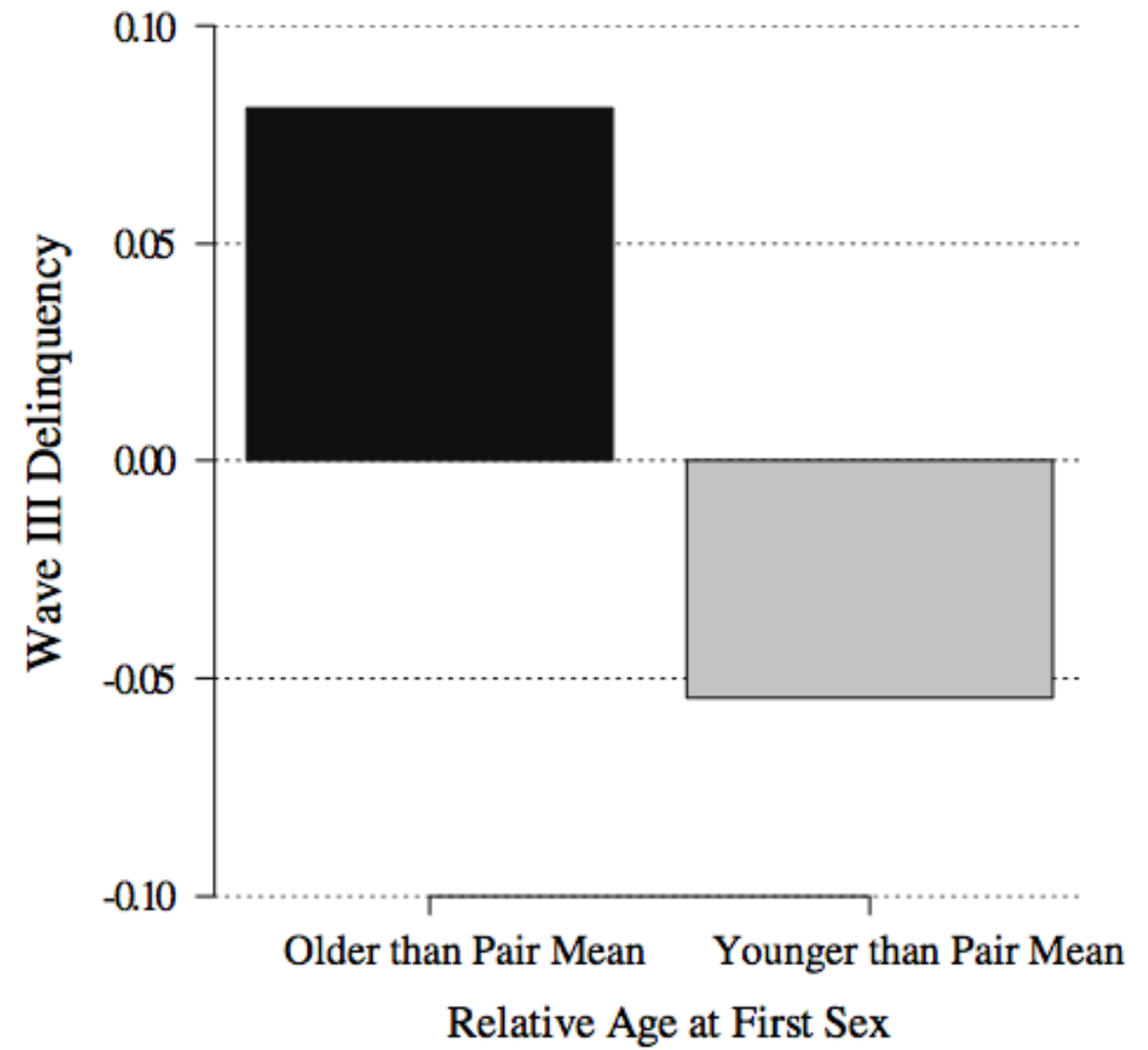
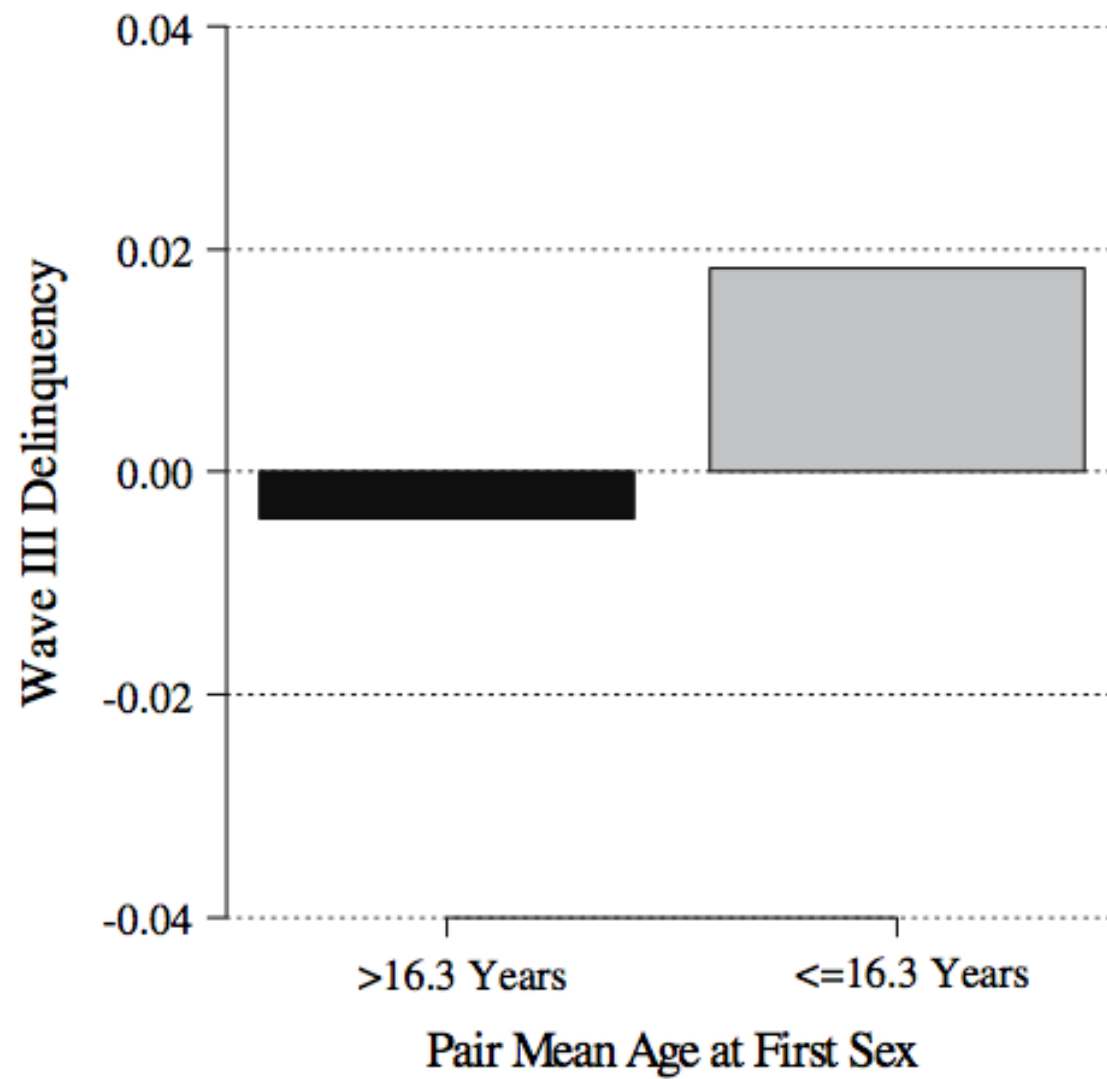
Quasi-experimental Designs

- Quasi-experimental designs (QEDs) are a family of methods for identifying naturally occurring circumstances in observational data that provide some of the power of controlled randomized experiments.
- QEDs are...
 - ...templates for causal inference that...
 - ...increase analytic power by...
 - ...exploiting structure in the data and existing causal knowledge in ways that...
 - ...emulate control and randomization

Many types of QEDs

- Twin design — Control the value of some potential common causes within specified pairs of instances.

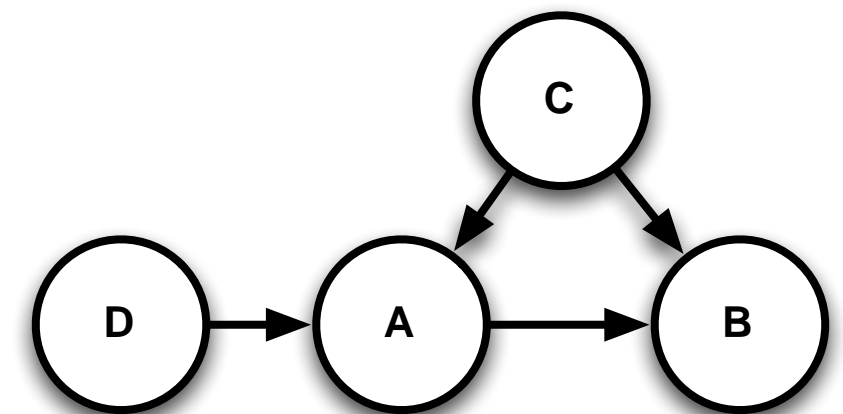
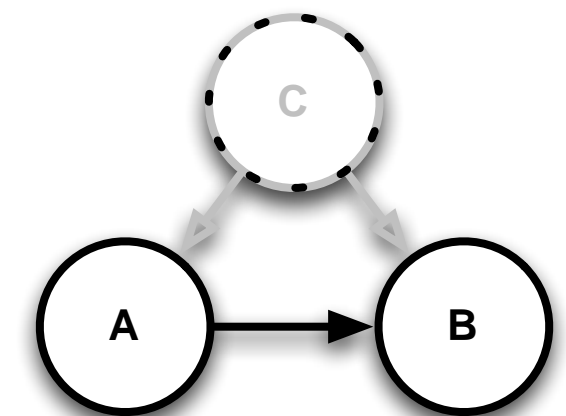
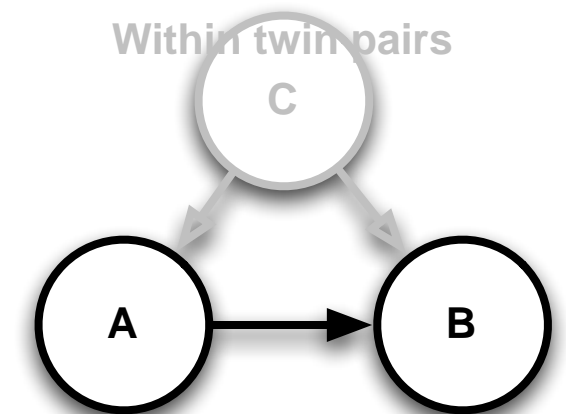




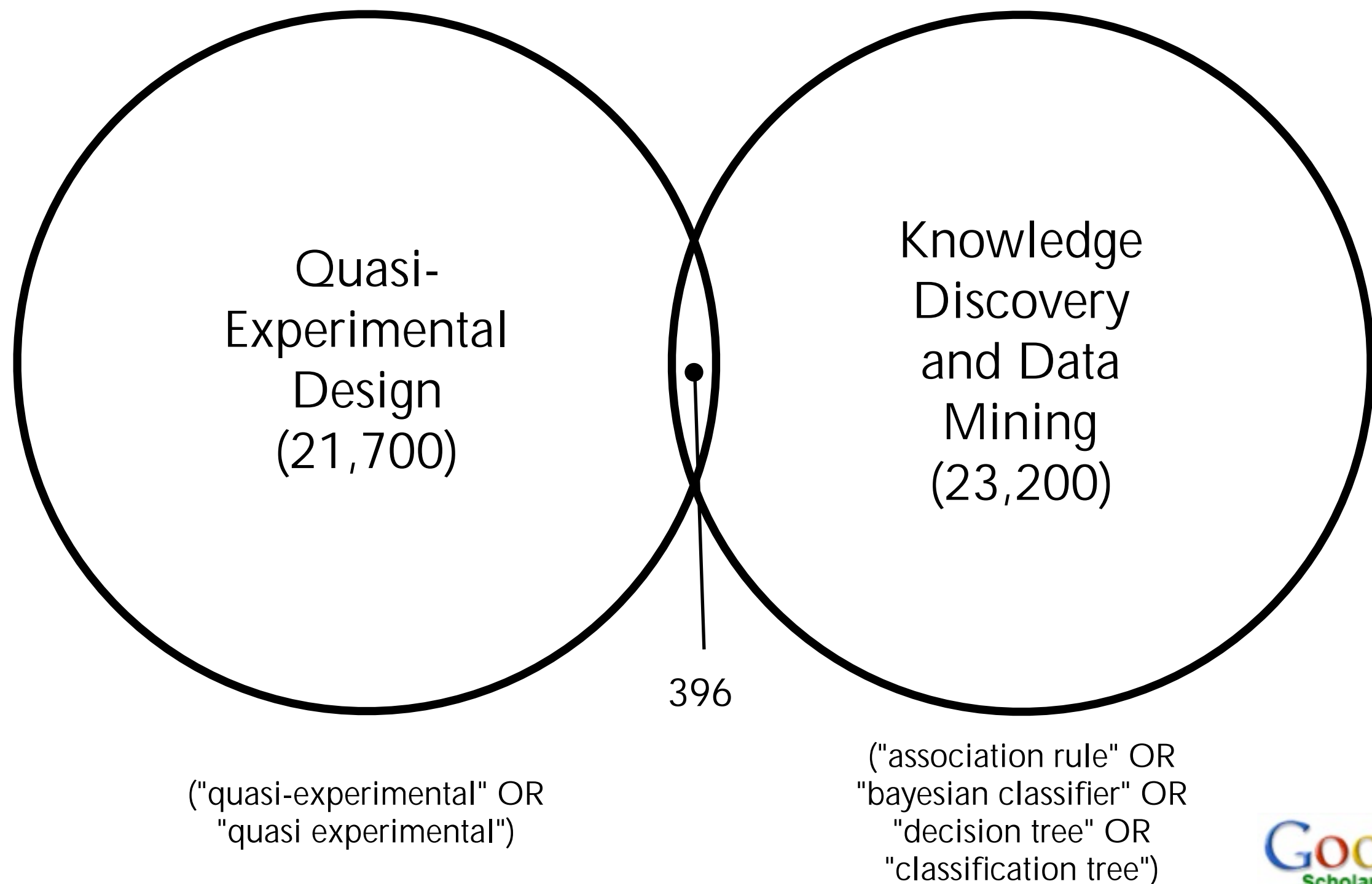
(Hardin et al. 2007)

Many types of QEDs

- Twin design — Control the value of some potential common causes within specified pairs of instances.
- Two-group pre-test post-test design — Compare temporal responses of treated instances to a control group of similar untreated instances.
- Instrumental variable design — Identify a variable that affects treatment, but is independent of all common causes.
- and many more...

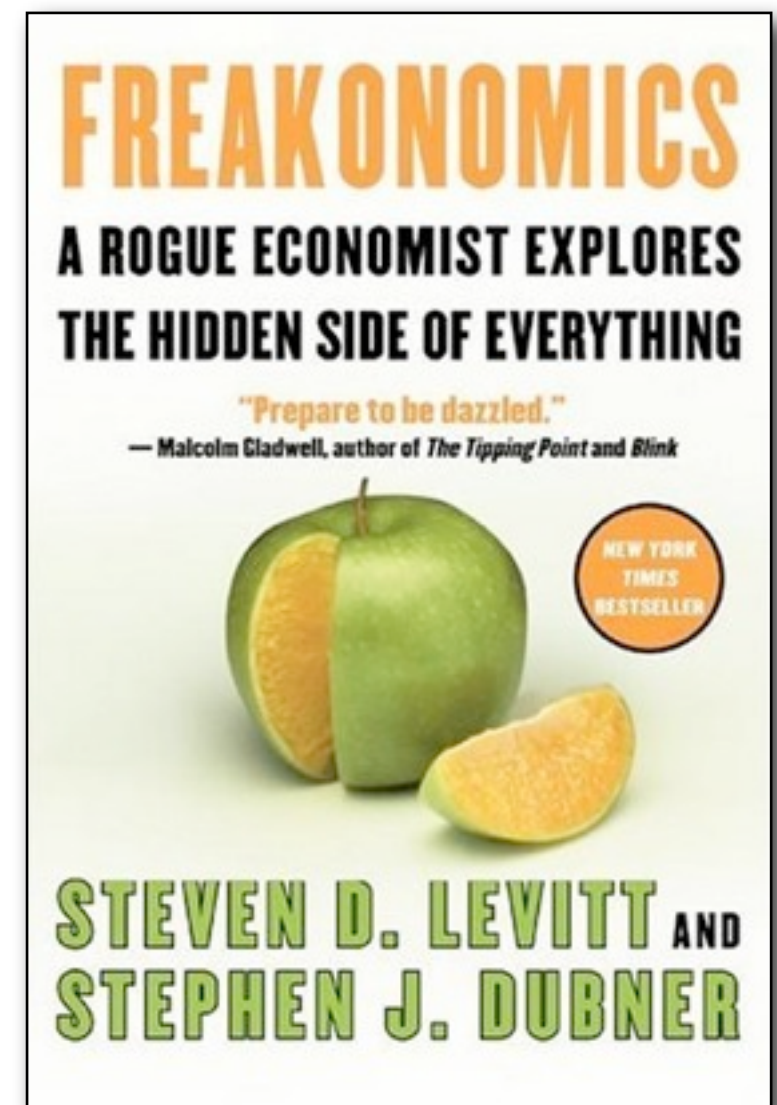


QEDs are frequently used



Finding QEDs can be difficult

- Identifying QEDs requires highly specific knowledge of both the domain and QED.
- We celebrate researchers who are creative and knowledgeable enough to find interesting and useful QEDs.
- All applications of QEDs to date have been entirely manual.

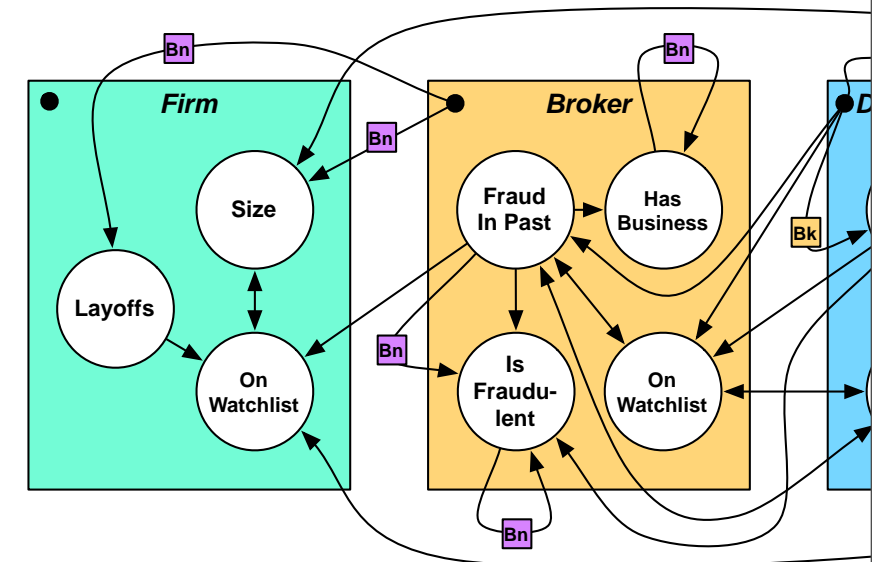


Assertion 4

Quasi-experimental designs
could enable a new generation
of algorithms and systems
for knowledge discovery

New opportunities for automation

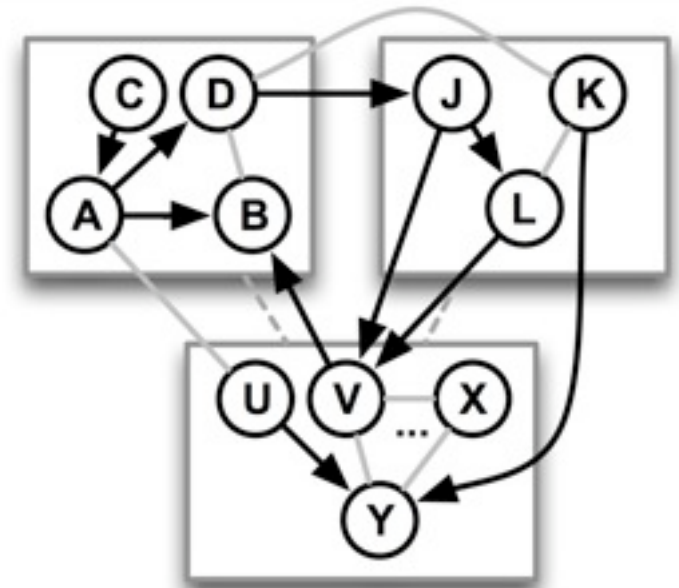
- Identifying QEDs requires representation and reasoning about relations and time.
- Recent developments make this possible
 - Widespread use of relational databases — Datasets with large and complex relational schemas are increasingly available
 - Development of relational models — New methods from relational learning and social network analysis



New challenges as well

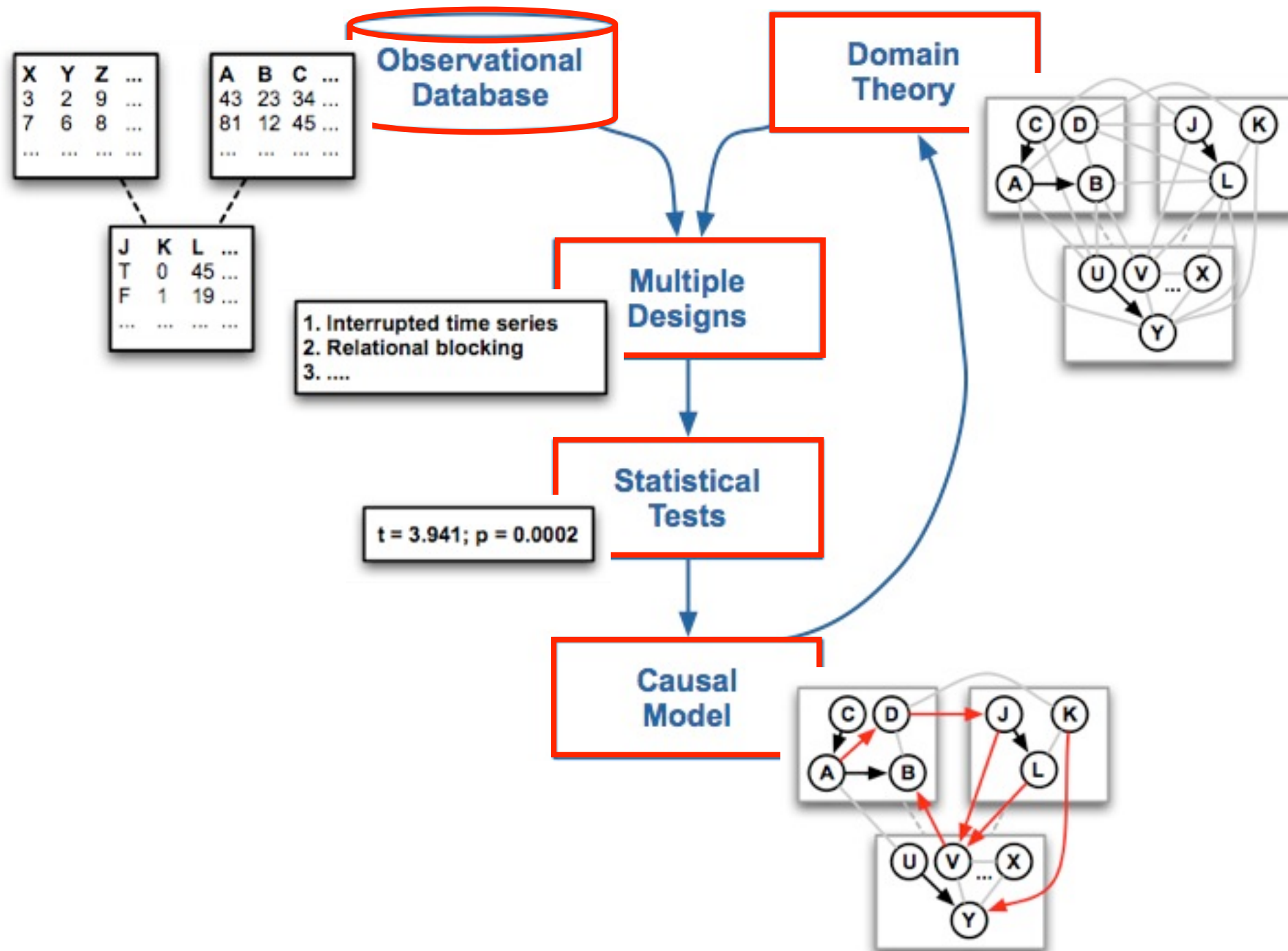
- Space of possible models is vast. For N variables, the number of possible causal models is:

$$3^{N^2} \approx 3^{(N^2)}$$



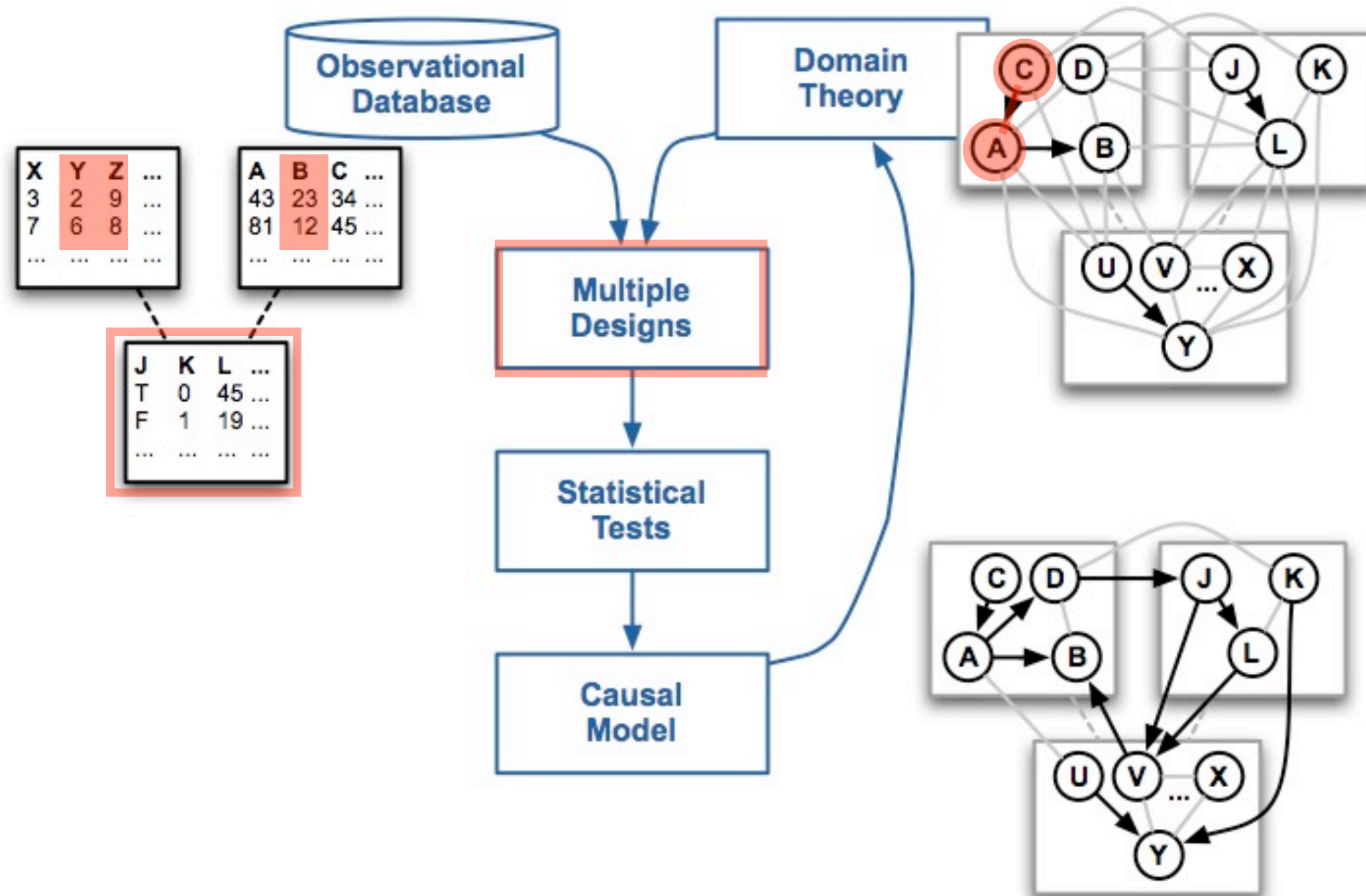
Variables	Possible causal models
3	27
6	14,348,907
9	150,094,635,296,999,000

Automatic application of QEDs



Quasi-experimental designs

QEDs are templates for reasoning that specify conditions in the data and domain theory that allow for valid causal inference.



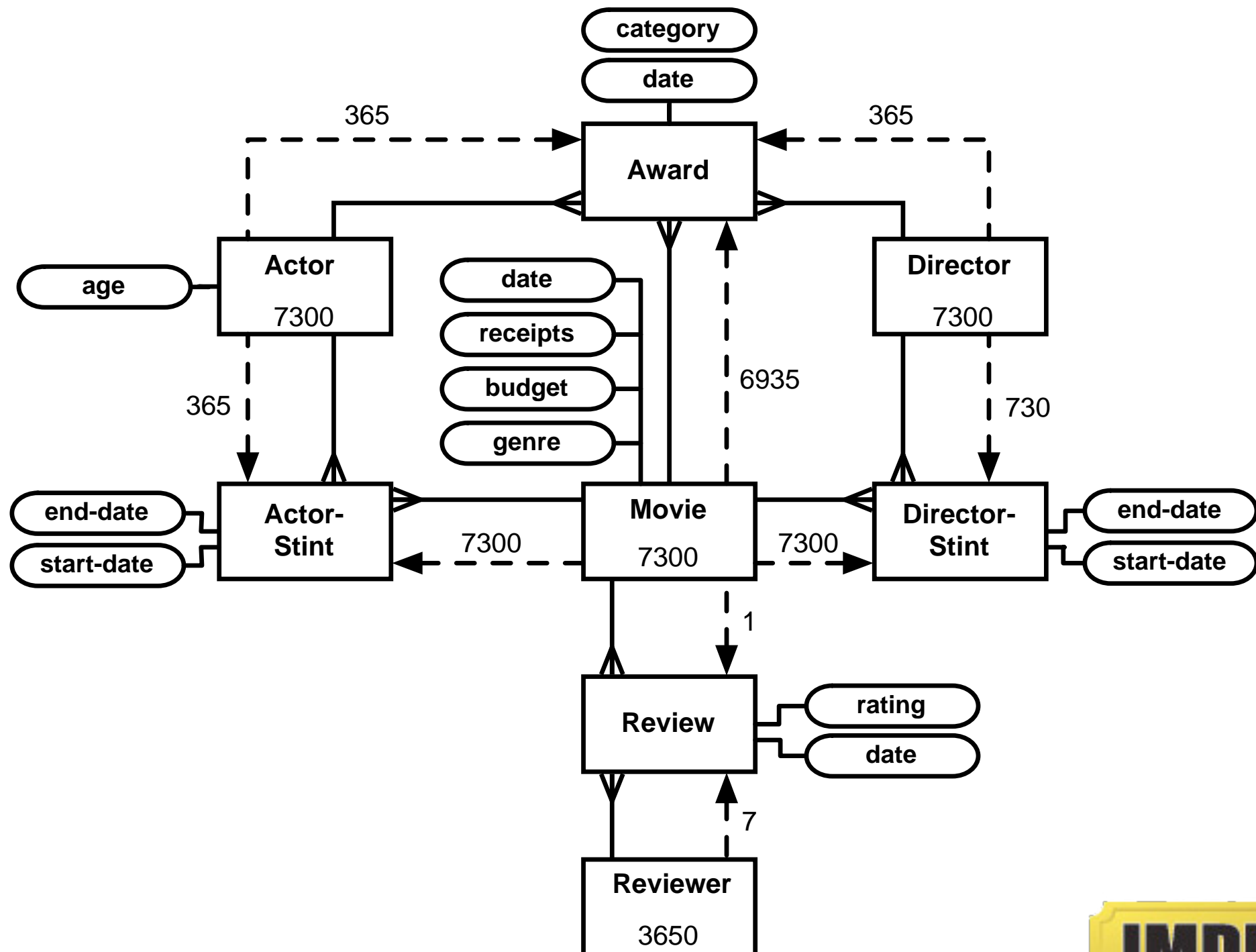
AIQ

- AIQ (Automated Identification of Quasi-experiments) is a prototype system for finding QEDs.
- Input — Standard ER diagram annotated with temporal extents and frequencies;
Domain knowledge about known causes
- Output — Specification of a QED, including treatments, outcomes, and units

AIQ

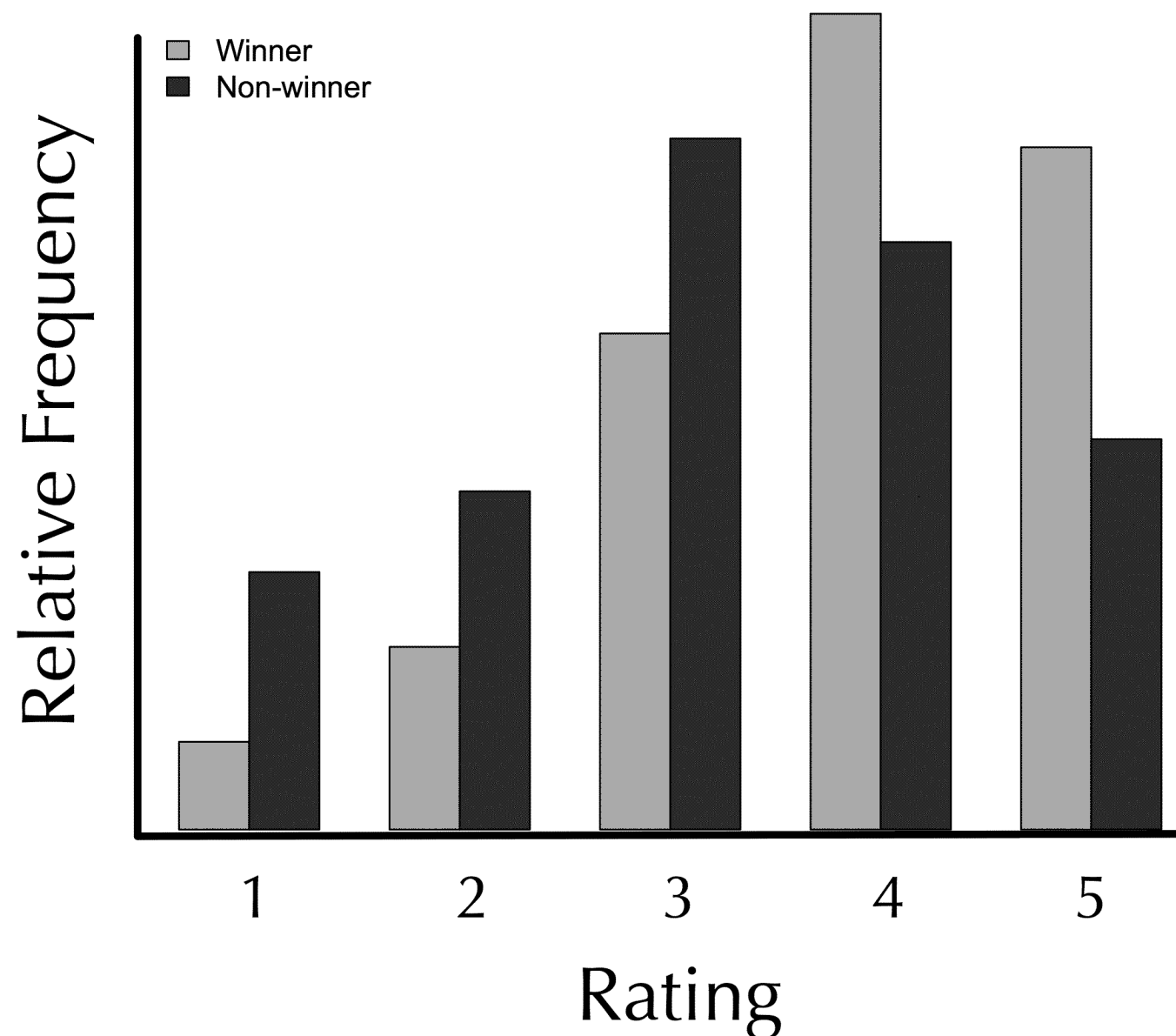
- AIQ augments the given data schema
 - Streams — Data structures created from periodic events (e.g., movie releases)
 - Aggregated variables — Variables created to measure aspects of streams (e.g., rate of movie releases)
- AIQ constructs units (data instances) that are joins of records
- AIQ matches constructed elements to QEDs

Example — Input



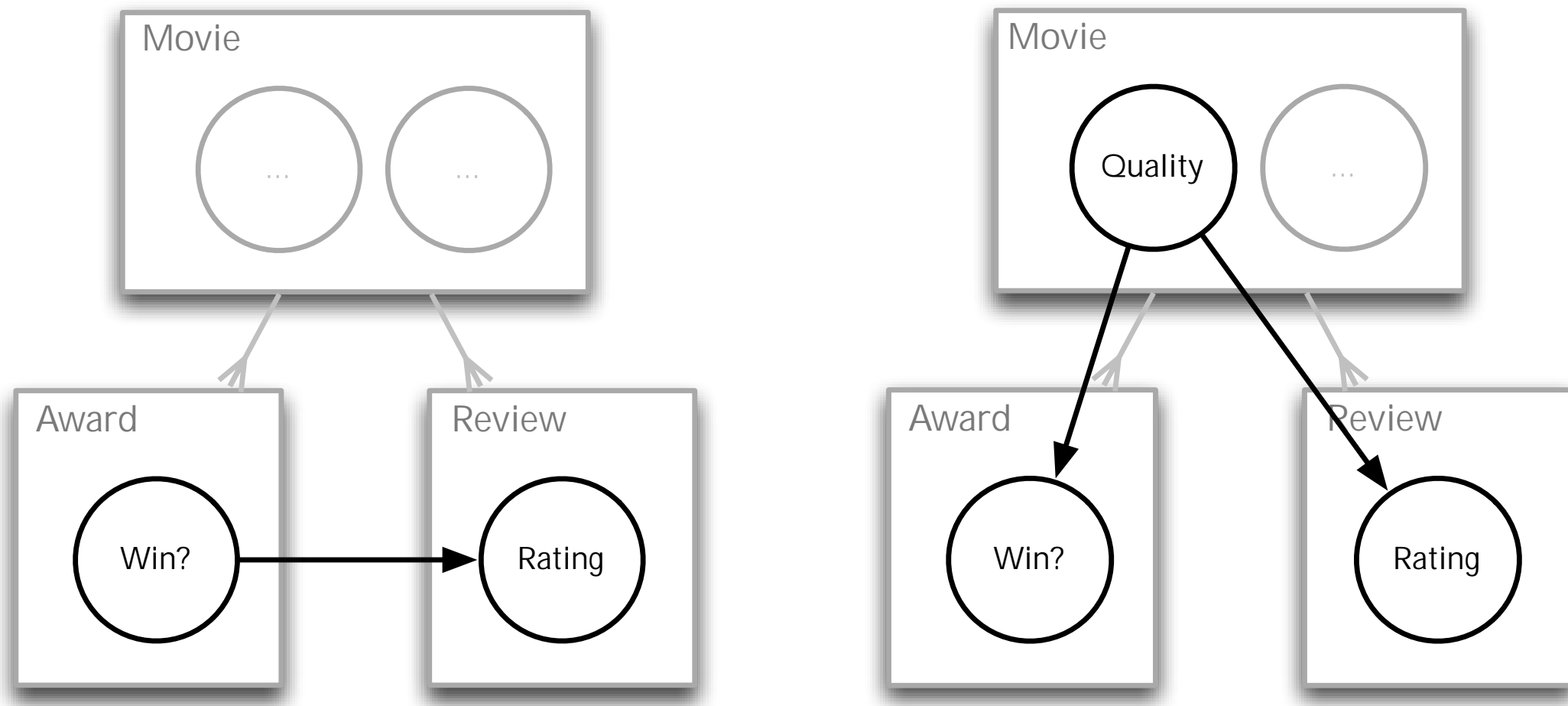
Causal question

Does winning an award cause a change in audience ratings of a movie?



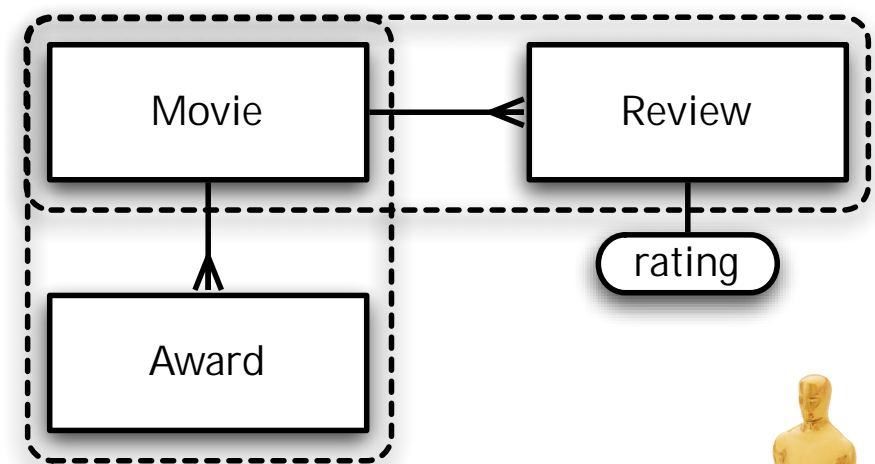
Causal question

Does winning an award cause a change in audience ratings of a movie?



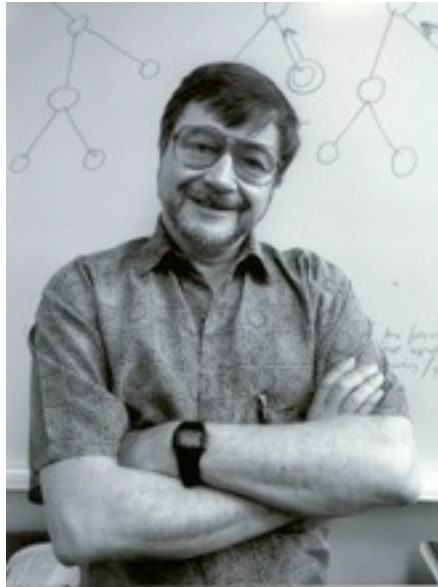
Example — Output

- Instance of non-equiv. control group design
- Units — Movies with associated awards and reviews
- Treatment — Winning an Academy Award
- Outcome — Netflix rating of movie
- Fortunately, awards can be treated as “quasi-random” within nominated movies



Example — Findings

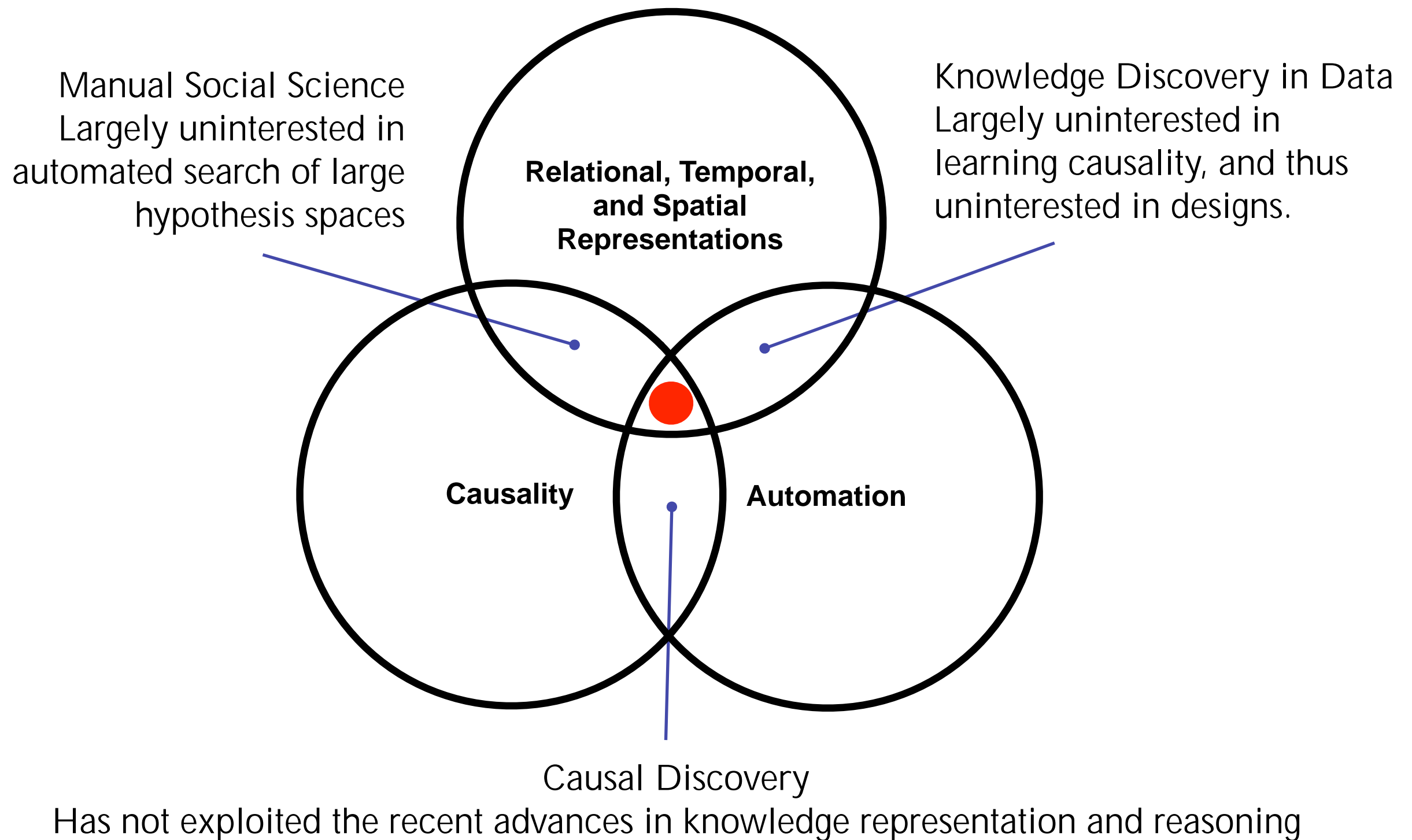
- We examined the difference between average ratings two months before and after award
- Among nominated movies, Netflix ratings decrease by...
 - 0.247 for non-winning movies
 - 0.066 for winning movies
- Difference is weakly significant ($p=0.07$), indicating a causal relationship



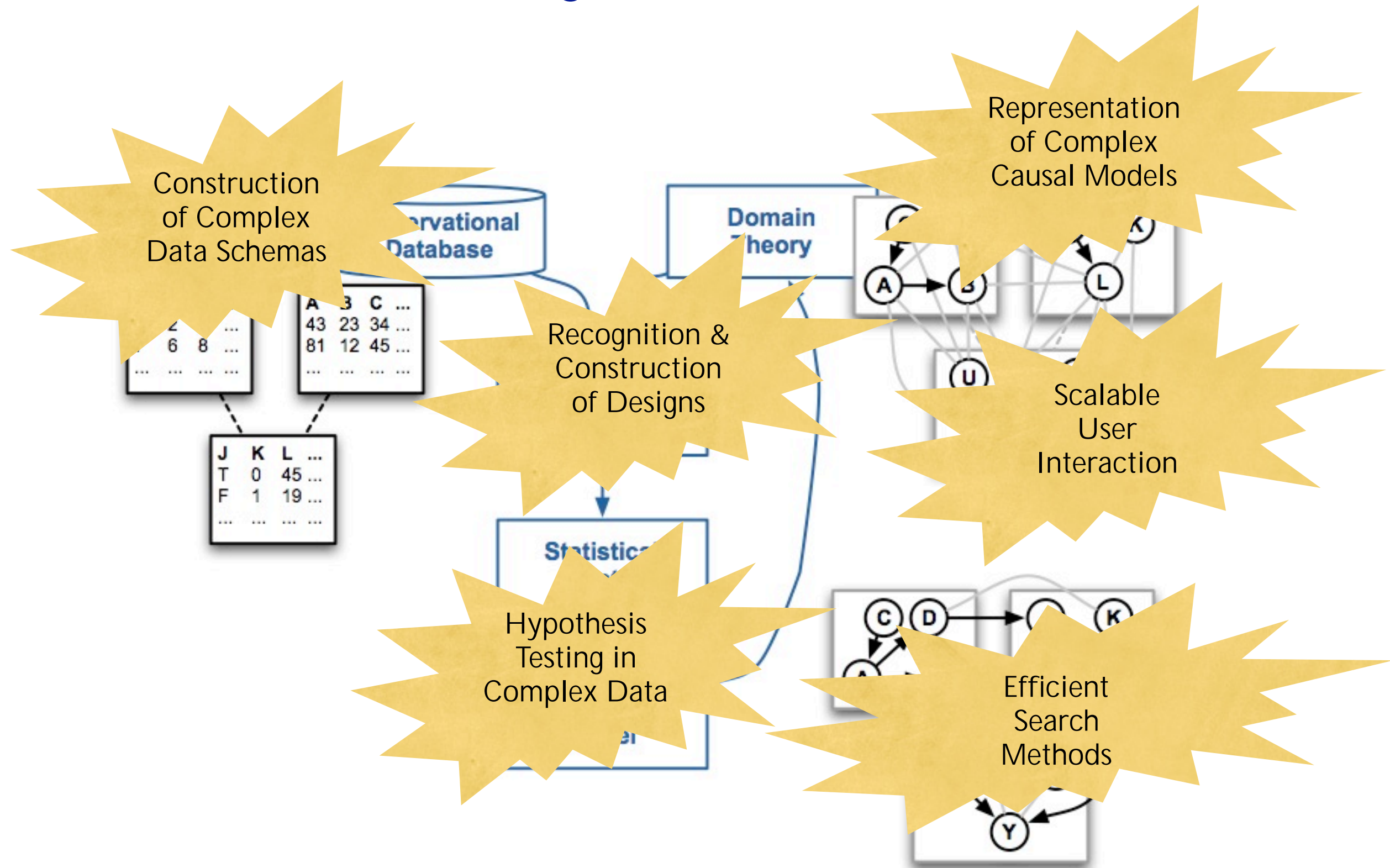
"...I see no greater impediment
to scientific progress
than the prevailing practice of focusing
all of our mathematical resources
on probabilistic and statistical inferences
while leaving causal considerations
to the mercy of intuition and good judgment."

- Judea Pearl (2000)

Existing research is fragmented



Several key research issues



Contact:

jensen@cs.umass.edu
kdl.cs.umass.edu
/papers, /proximity

Thanks to:

Brian Taylor, Andrew Fast, Lisa Friedland,
Henry Goldberg, Michael Hay, John Komoroske,
Marc Maier, Jennifer Neville, Matthew Rattigan,
Ted Senator, and Agustin Schapira

