UNIVERSITÀ DEGLI STUDI DI TORINO
ALMA UNIVERSITAS TAURINENSIS

Some comments on Reproducibility, Repeatability, Replicability and Generalizability

Professor Ron S. Kenett

Samuel Neaman Institute for National Policy Research

KPA Insights through analytics

Reproducible Research

Marcia McNutt is Editor-in-Chief of *Science*.

Reproducibility

SCIENCE ADVANCES ON A FOUNDATION OF TRUSTED DISCOVERIES. REPRODUCING AN EXPERIMENT is one important approach that scientists use to gain confidence in their conclusions. Recently, the scientific community was shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible. Because confidence in results is of paramount importance to the broad scientific community, we are announcing new initiatives

www.sciencemag.org **SCIENCE** VOL 343 17 JANUARY 2014

Published by AAAS

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands

KPA Insights through analytics

Replicable Research

The Economist

Problems with scientific research

How science goes wrong

Scientific research has changed the world. Now it needs to change itself

Oct 19th 2013 | From the print edition

NEWS FEATURE

NATURE|Vol 442|27 July 2006

THE TROUBLE WITH REPLICATION

The idea that readers should be able to replicate published scientific results is seen as the bedrock of modern science. But what if replication proves difficult or impossible? **Jim Giles** tracks the fate of one group of papers.

KPA Insights through analytics

January 10th, 2017

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

replicable

A manifesto for ~~reproducible~~ science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

Improving the reliability and efficiency of scientific research will increase the credibility of the published scientific literature and accelerate discovery. Here we argue for the adoption of measures to optimize key elements of the scientific process: methods, reporting and dissemination, reproducibility, evaluation and incentives. There is some evidence from both simulations and empirical studies supporting the likely effectiveness of these measures, but their broad adoption by researchers, institutions, funders and journals will require iterative evaluation and improvement. We discuss the goals of these measures, and how they can be implemented, in the hope that this will facilitate action toward improving the transparency, reproducibility and efficiency of scientific research.

KPA Insights through analytics

“I placed too much faith in underpowered studies:” Nobel Prize winner admits mistakes

February 20th, 2017

with 6 comments

Although it’s the right thing to do, it’s never easy to admit error — particularly when you’re an extremely high-profile scientist whose work is being dissected publicly. So while it’s not a retraction, we thought this was worth noting: A Nobel Prize-winning researcher has admitted on a blog that he relied on weak studies in a chapter of his bestselling book.

Daniel Kahneman

The blog — by Ulrich Schimmack, Moritz Heene, and Kamini Kesavan — critiqued the citations included in a book by Daniel Kahneman, a psychologist whose research has illuminated our understanding of how humans form judgments and make decisions and earned him half of the 2002 Nobel Prize in Economics.

According to the Schimmack et al blog,

> …readers of his [Kahneman’s] book “Thinking Fast and Slow” should not consider the presented studies as scientific evidence that subtle cues in their environment can have strong effects on their behavior outside their awareness.

Priming (psychology)

From Wikipedia, the free encyclopedia

Priming is an implicit memory effect in which exposure to one stimulus (i.e., perceptual pattern) influences the response to another stimulus. The seminal experiments of Meyer and Schvaneveldt in the early 1970s[1][2][3] led to the flowering of research on priming of many sorts. Their original work showed that people were faster in deciding that a string of letters is a word when the word followed an associatively or semantically related word. For example, NURSE is recognized more quickly following DOCTOR than following BREAD. Various experiments[2][3] supported the theory that activation spreading among related ideas was the best explanation for the facilitation observed in the lexical decision task. The priming paradigm provides excellent control over the effects of individual stimuli on cognitive processing and associated behavior because the same target stimuli can be presented with different primes. Thus differences in performance as a function of differences in priming stimuli must be attributed to the effect of the prime on the processing of the target stimulus.

KPA
Insights through analytics

Reproducibility versus Replicability

Replicability is not Reproducibility: Nor is it Good Science

Chris Drummond
Institute for Information Technology
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6

CHRIS.DRUMMOND@NRC-CNRC.GC.CA

Proc. of the Evaluation Methods for Machine Learning Workshop at the 26 th ICML, Montreal, Canada, 2009.

"Reproducibility requires changes; replicability avoids them. A critical point of reproducing an experimental result is that irrelevant things are intentionally not replicated. One might say, one should replicate the result **not the experiment**."

A highly standardized experiment supplies direct information only in respect of the narrow range of conditions achieved by standardization. Standardization, therefore, weakens rather than strengthens our ground for inferring a result, when, as is the case in practice, these conditions are somewhat varied.

Ronald A. Fisher 1935

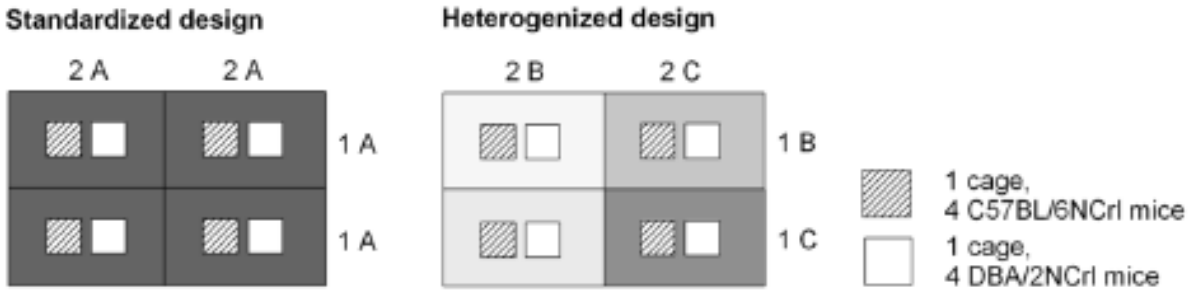
Kenett, R.S. and Shmueli, G. (2015) Clarifying the terminology that describes scientific reproducibility, Nature Methods, Vol. 12(8), p 699.

KPA Insights through analytics

Reproducibility in Animal Behavior

- Standardization is the attempt to increase reproducibility at the expense of external validity
- Standardization reduces external validity and thus also reproducibility
- Heterogenization increases external validity and thus also reproducibility

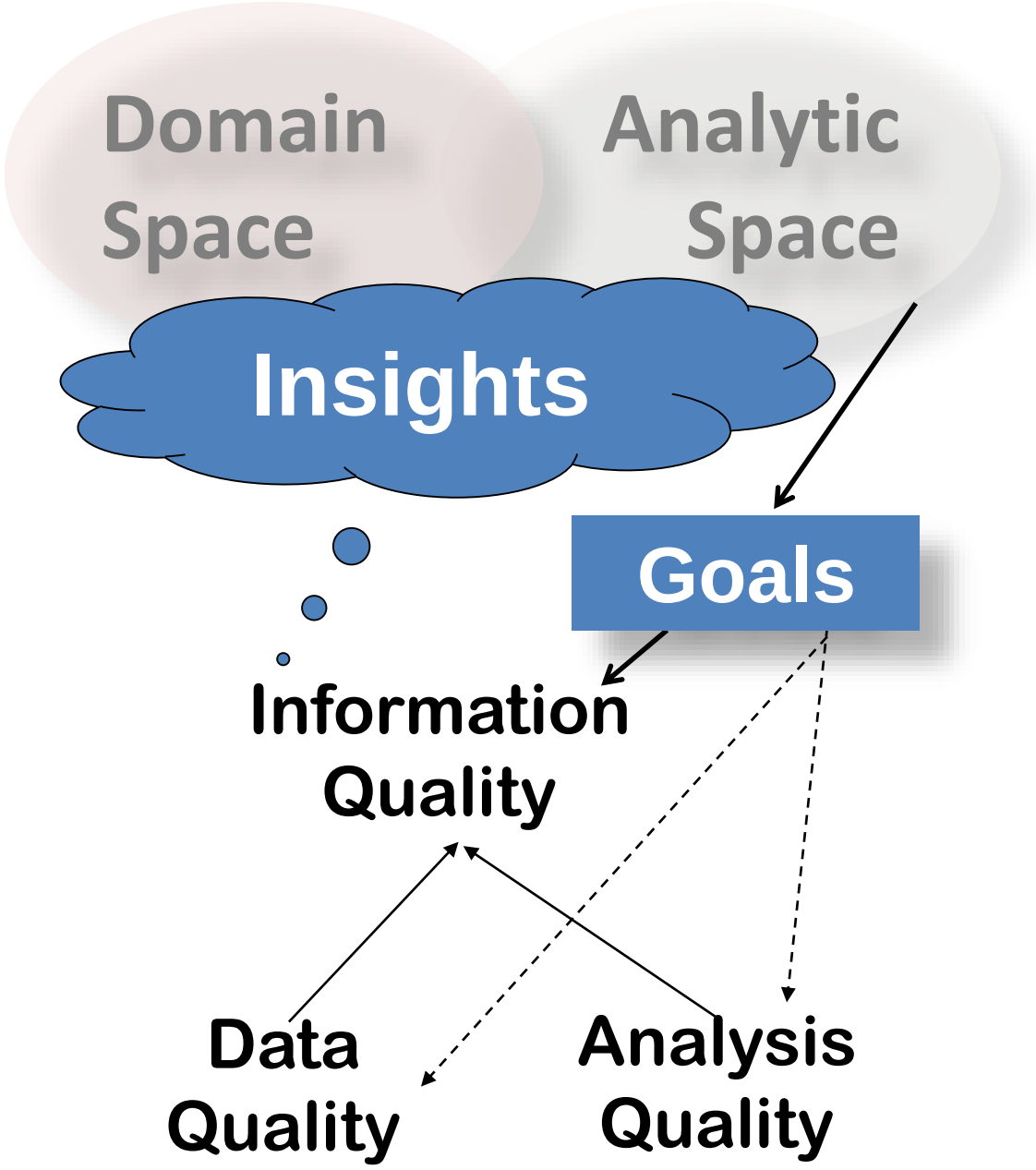
Würbel et al. 2000 Nature Genetics
Richter et al. 2010 Nature Methods
Richter et al. 2011 PLoS ONE



Experimental factors		Factor level A	Factor level B	Factor level C
1	Test age of the animals	12 weeks old	8 weeks old	16 weeks old
2	Cage enrichment	Nesting material	Shelter (MouseHouse), nesting material	Climbing structures, nesting material

KPA
Insights through analytics

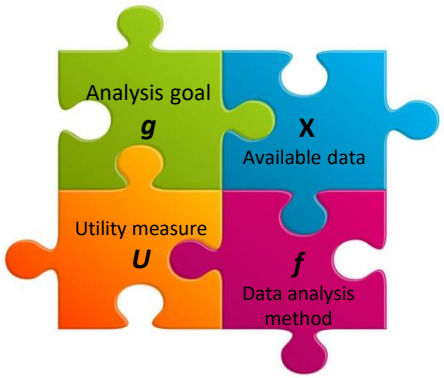
$$InfoQ(f,X,g) = U(f(X|g))$$



InfoQ Dimensions

1. Data resolution
2. Data structure
3. Data integration
4. Temporal relevance
5. Chronology of data and goal
6. **Generalizability**
7. Operationalization
8. Communication

How

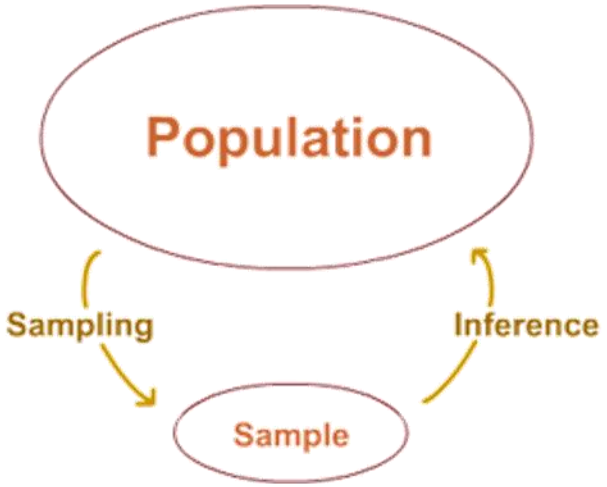


What

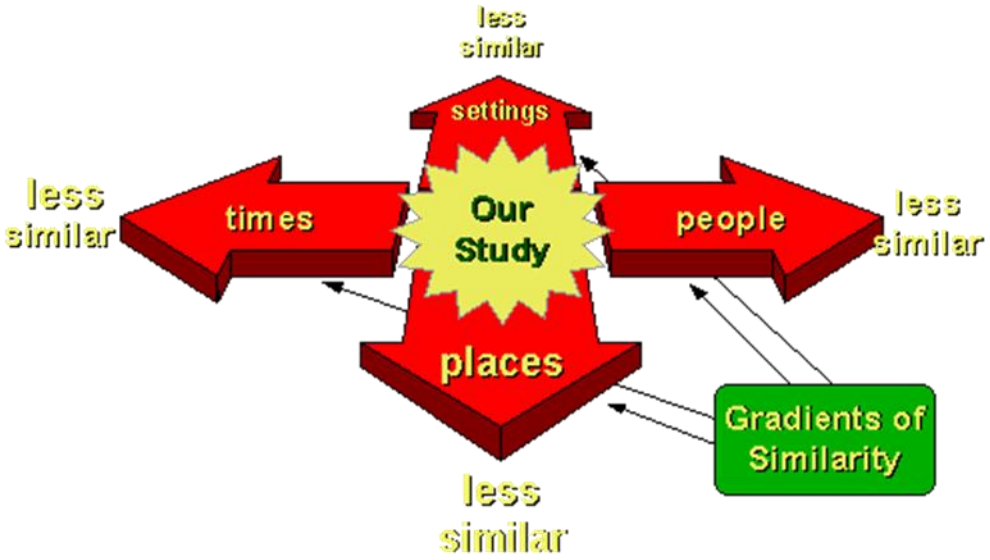
KPA
Insights through analytics

Generalizability

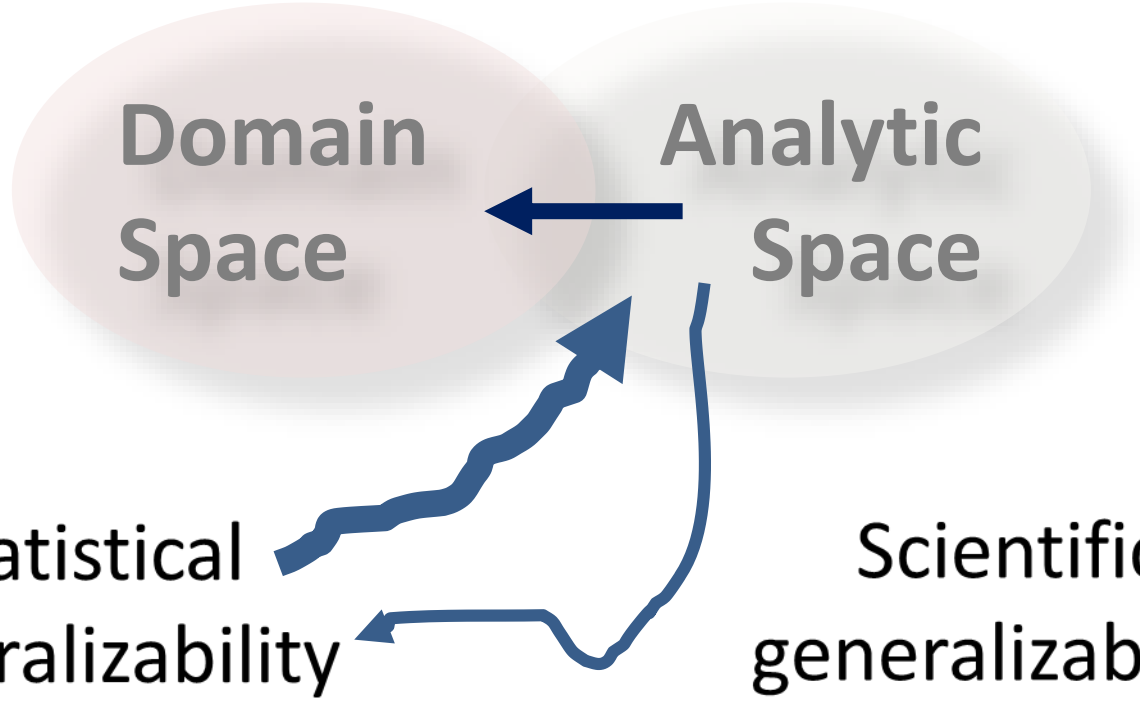
Statistical generalizability

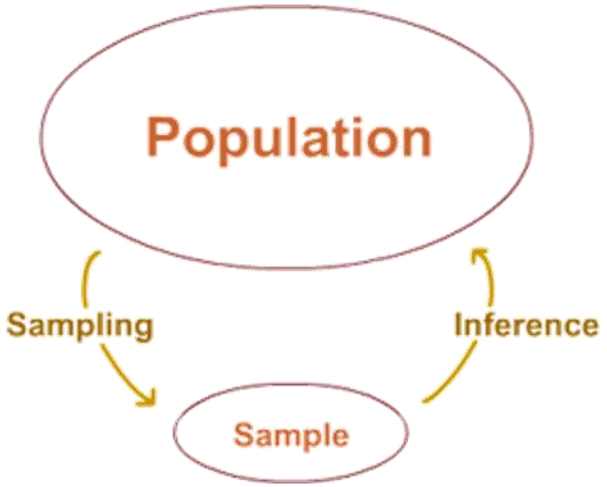


Scientific generalizability



KPA
Insights through analytics





KPA
Insights through analytics

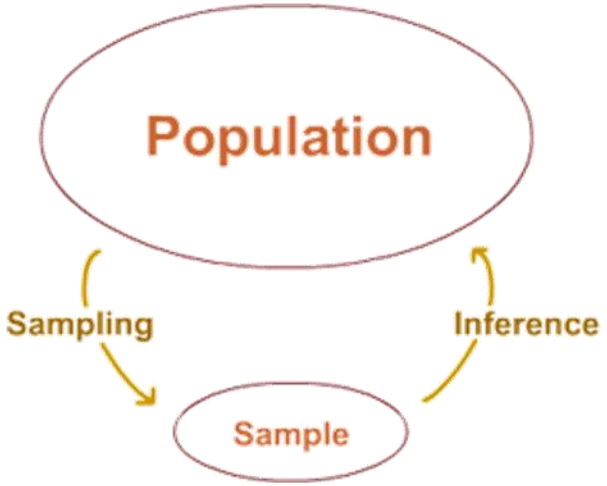
Domain Space

Analytic Space

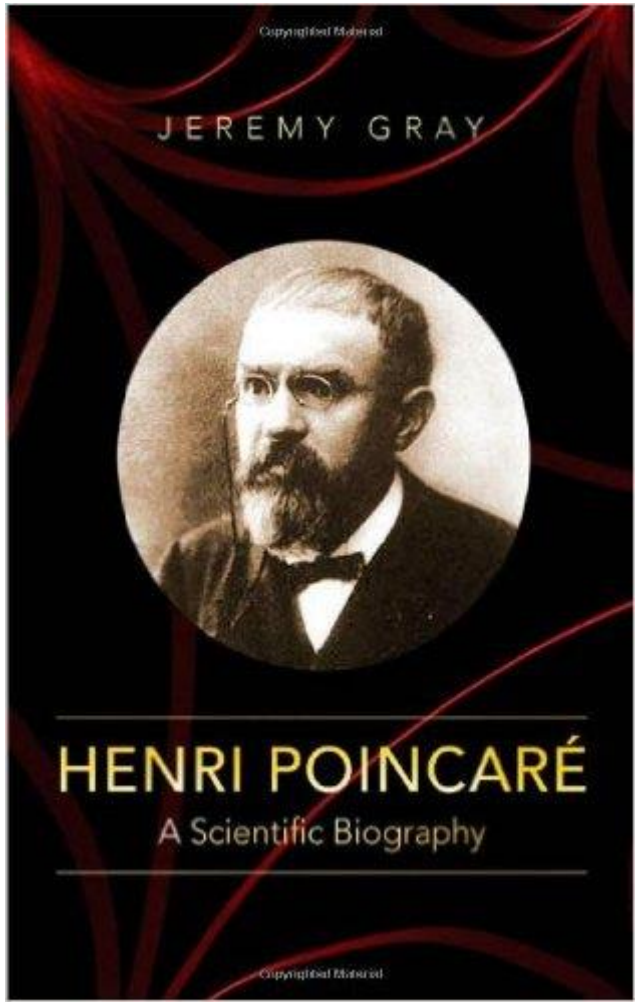
Conservation of Mass
Conservation of Energy
Conservation of Momentum
Newton Laws
Principle of least action
Laws of thermodynamics
Maxwell's equations
…..

Statistical generalizability

Scientific generalizability



KPA
Insights through analytics



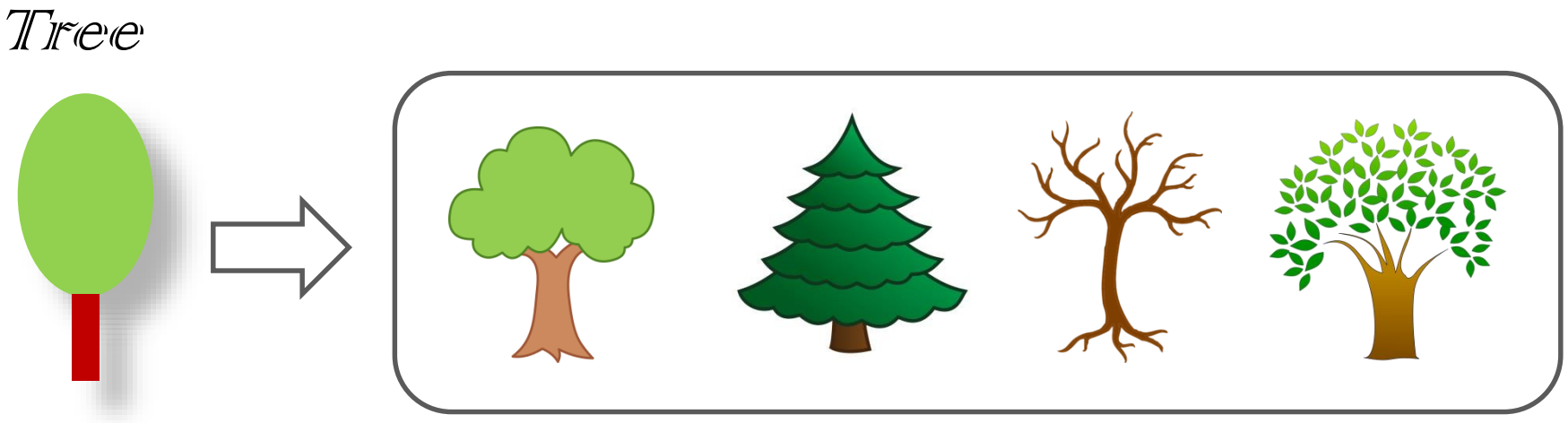
Princeton University Press, 2012

*"What he emphasized above all was the act of human **understanding**. His preferred means of attaining the understanding of a problem was to find the right **generalization** of its **core concepts**, often in the form of an **analogy**."*

J. Gray, preface to Henri Poincare, a scientific biography

"A **concept** is an abstraction or generalization from experience or the result of a transformation of existing concepts."

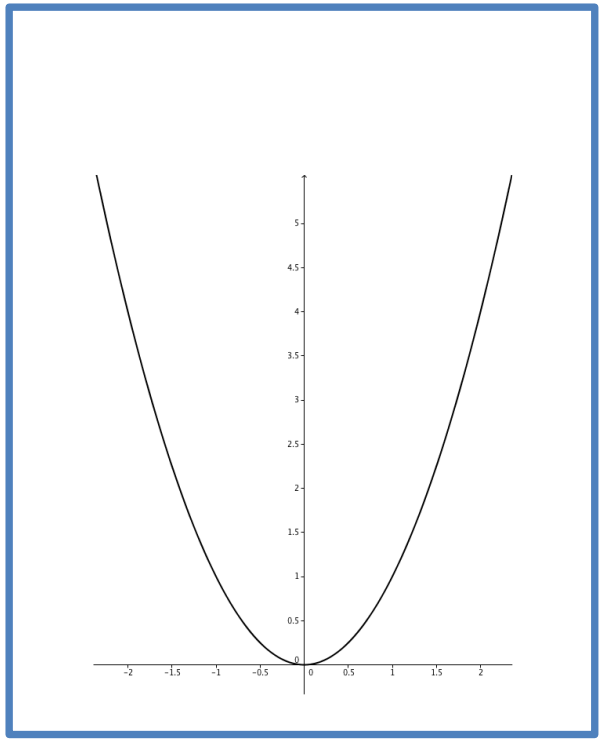
Wikipedia



A concept can be represented in alternative forms

KPA
Insights through analytics

Alternative representations with Meaning Equivalence

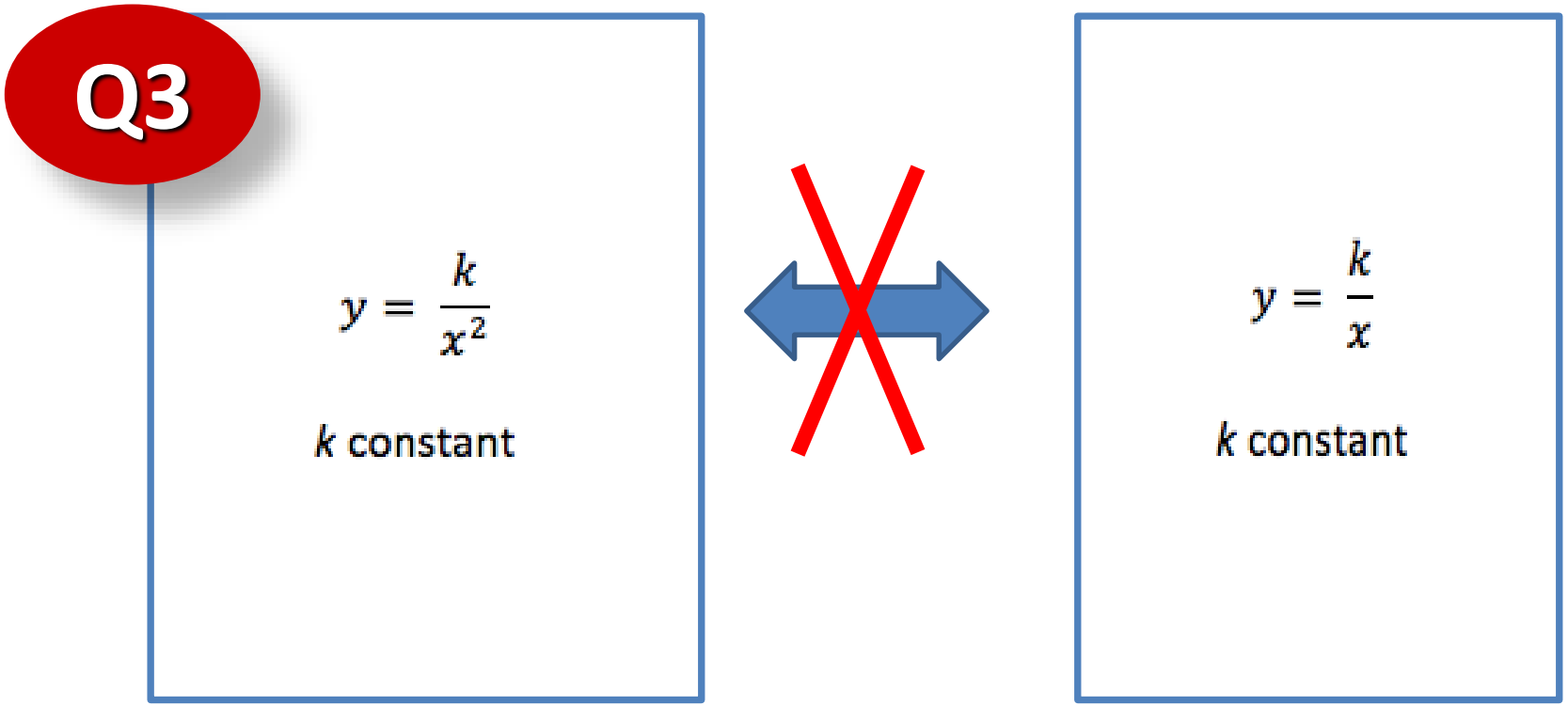


Q2

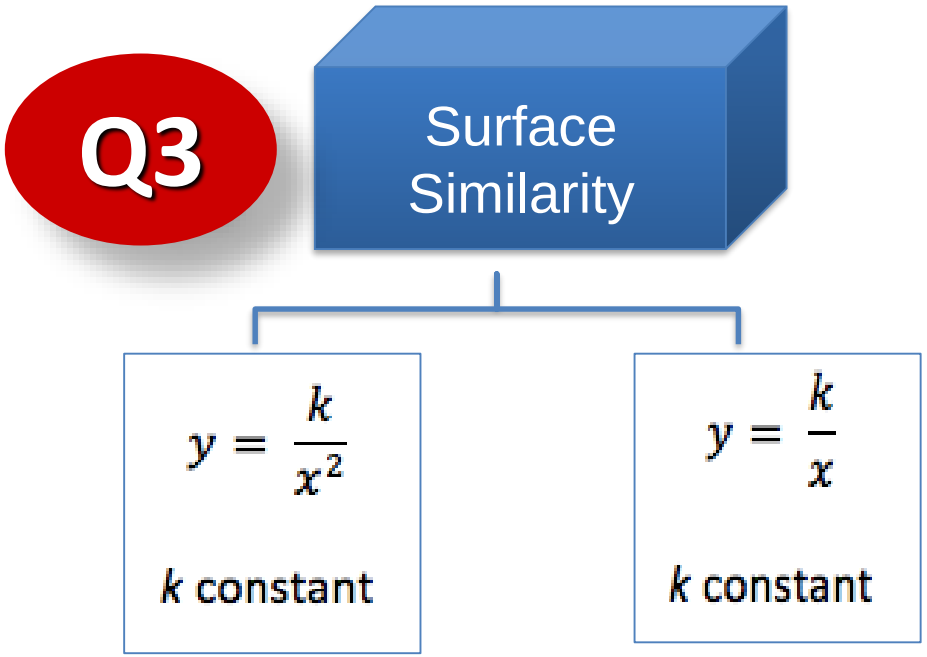
$Y = X^2$

KPA
Insights through analytics

Alternative representations with Surface Similarity



Alternative representations with Surface Similarity



KPA
Insights through analytics

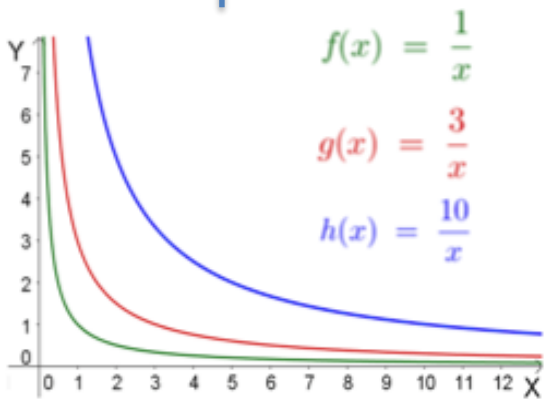
Alternative representations with Meaning Equivalence

Meaning Equivalence

Q2

$$y = \frac{k}{x}$$

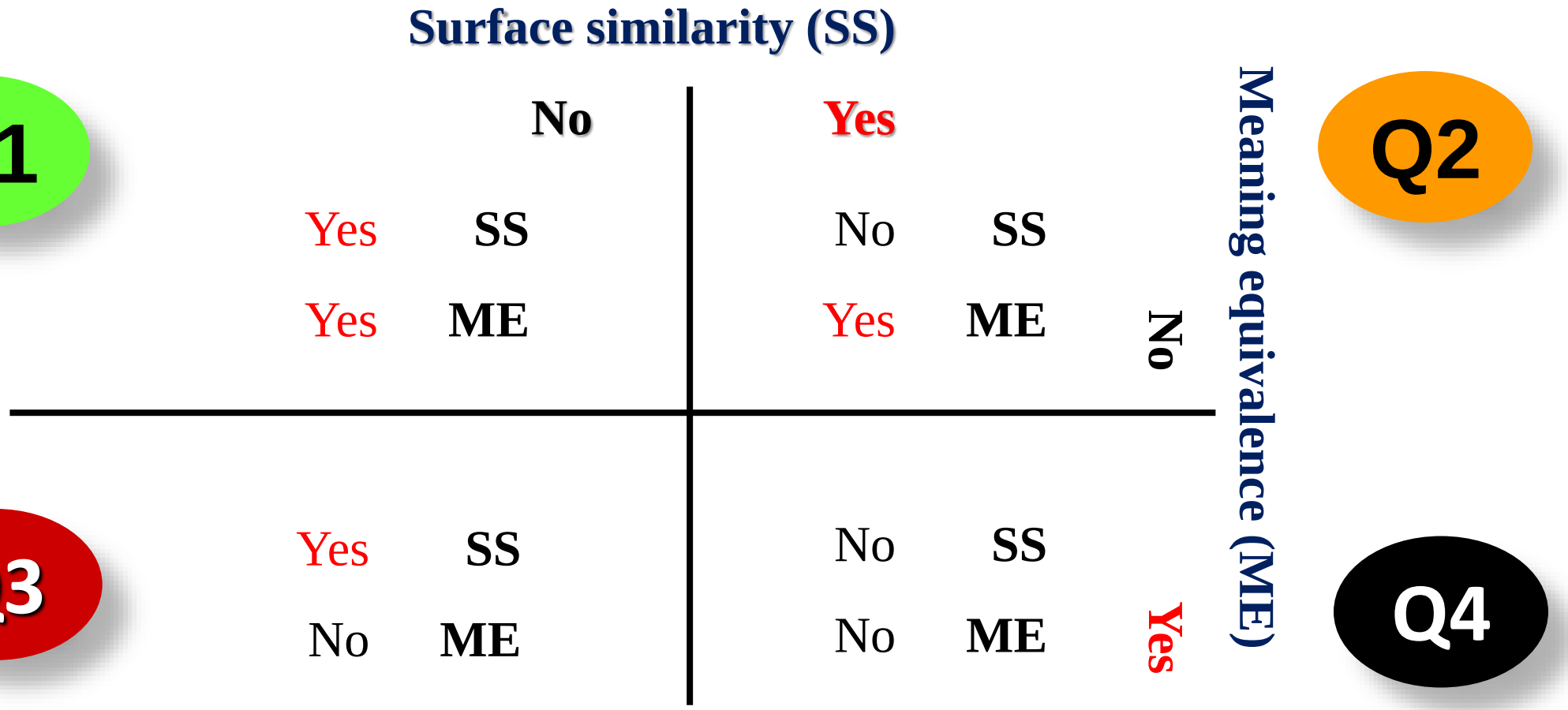
k constant



KPA Insights through analytics

Research findings

Generalize with Alternative Representations

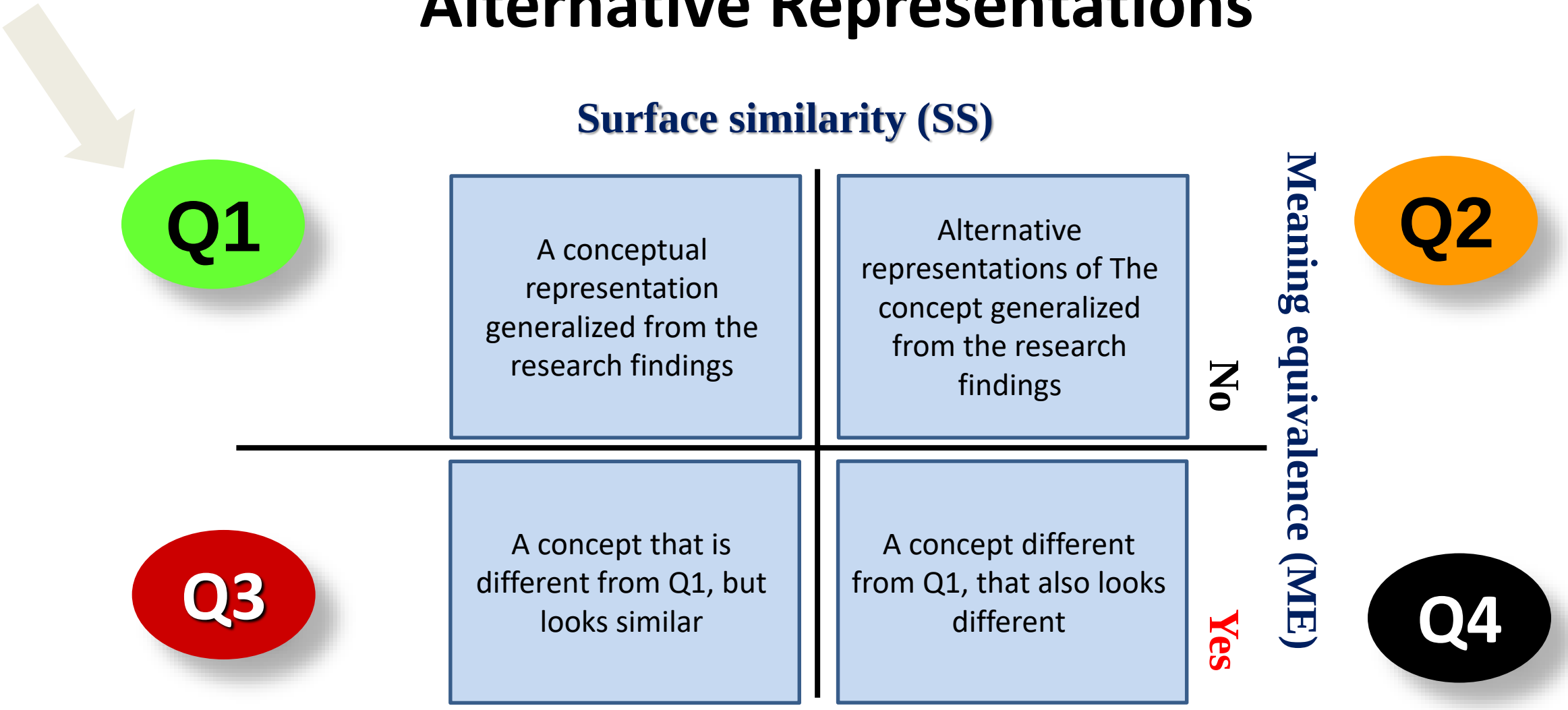


Shafrir, U. and Kenett, R.S. (2015), Concept Science Evidence-Based MERLO Learning Analytics, in Handbook of Applied Learning Theory and Design in Modern Education, IGI Global

KPA Insights through analytics

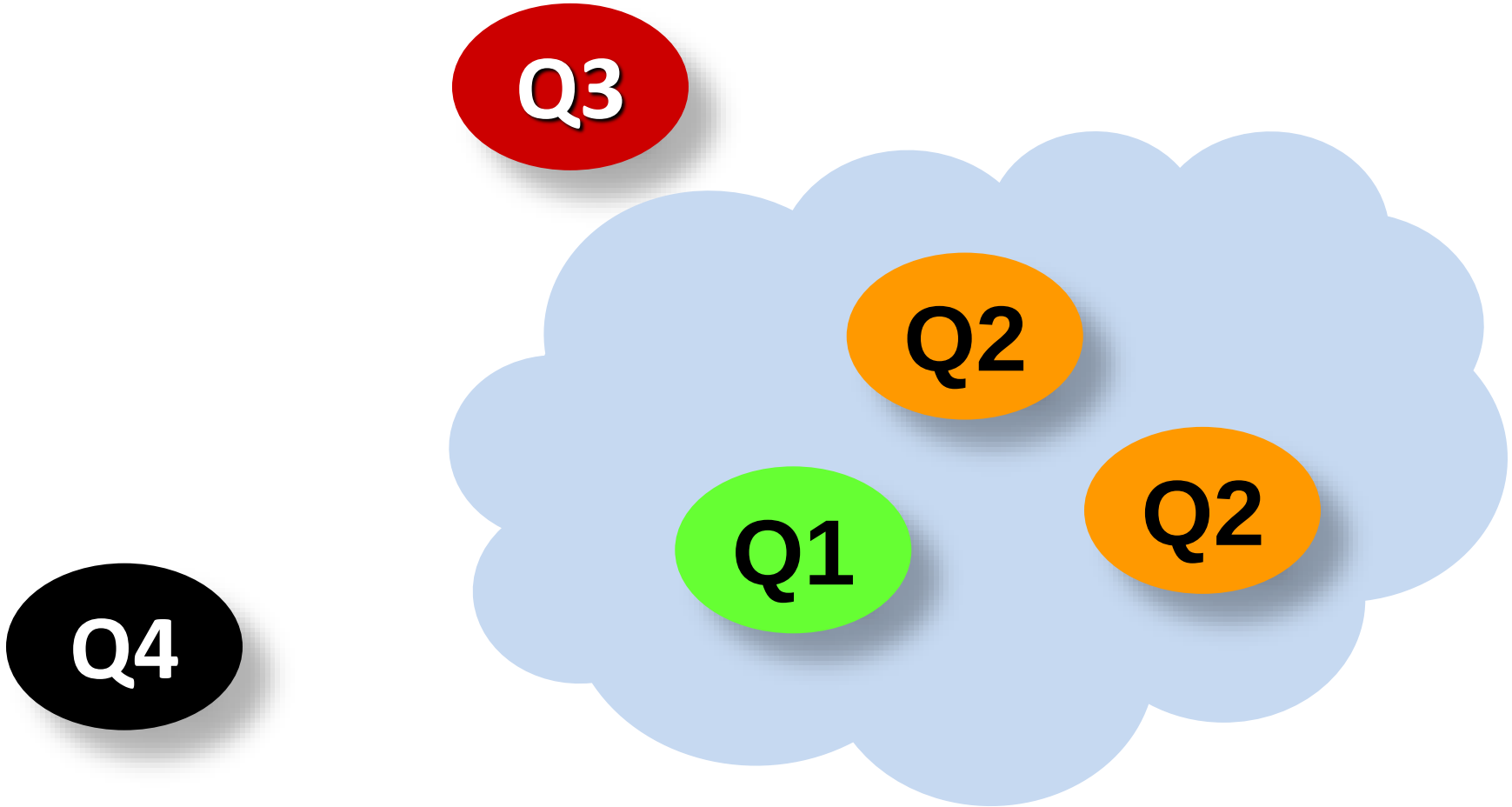
Research findings

Generalize with Alternative Representations



KPA Insights through analytics

Boundary of Meaning



Statistical Errors

A **Type I error** consists of rejecting the "null hypothesis" (roughly speaking, the assumption of no effect, the hypothesis you typically set out to dispro
in fact the null h

Type S error: I state that increasing X, increases Y and I am wrong

A **Type II error** c
failing to reject the null hypothesis) when in fact the null hypothesis is false.

Type S errors (sign errors, concluding : $\theta_1 > \theta_2$ when $\theta_1 < \theta_2$).

Type M errors (magnitude errors, concluding that an effect is larger than it truly is).

Gelman A. and Carlin, J. (2014), Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors, *Perspectives on Psychological Science*, Vol. 9(6), pp. 641–651.

From external information…
D : the true effect size

From the data (or model if prospective design)…
d : the observed effect
s : SE of the observed effect
p : the resulting p-value

Study Design

Hypothetical replicated data
d^{rep} : the effect that would be observed in a hypothetical replication study with a design like the one used in the original study (so assumed also to have SE = *s*)

This is essentially a Bayesian argument

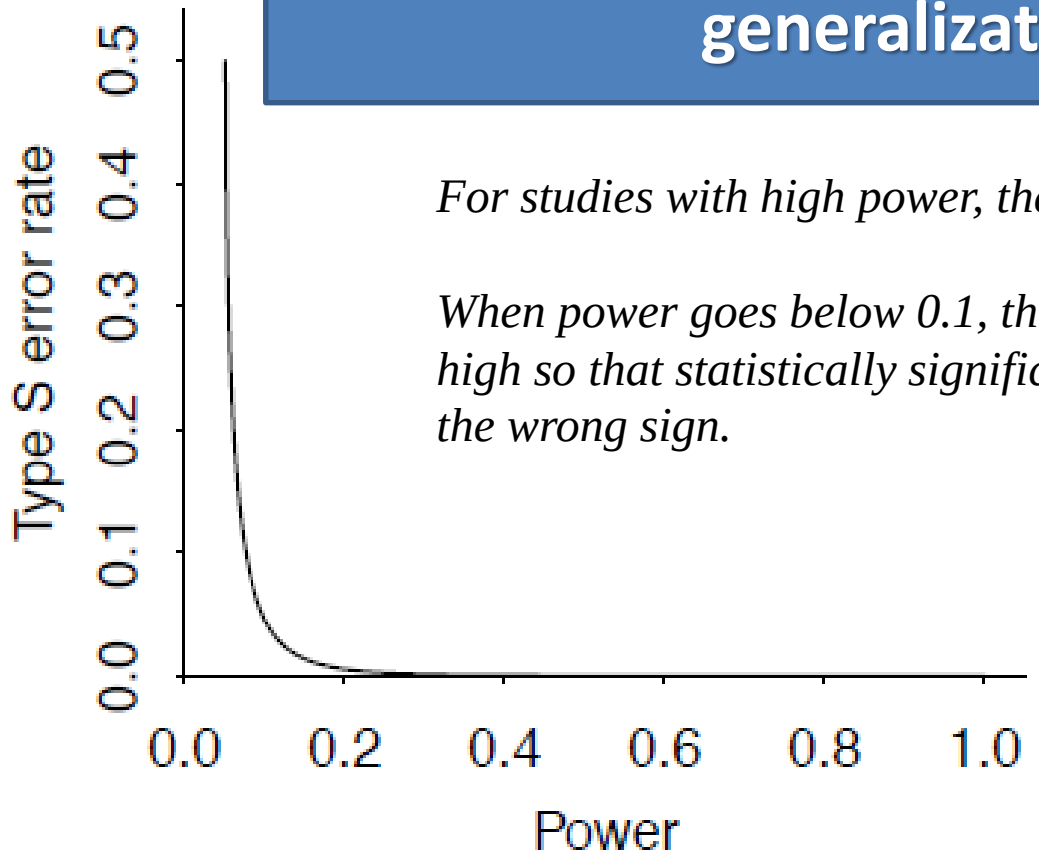
Design calculations:

- *Power*: the probability that the replication d^{rep} is larger (in absolute value) than the critical value that is considered to define "statistical significance" in this analysis.
- *Type S error rate*: the probability that the replicated estimate has the incorrect sign, if it is statistically significantly different from zero.

Type S error: $\theta 1 > \theta 2$, but I claim that $\theta 1 < \theta 2$ (or vice versa)

Type S (sign) errors

Probability (an error in scientific generalization)



For studies with high power, the Type S error rate is low.

When power goes below 0.1, the Type S error rate becomes high so that statistically significant estimates are likely to be the wrong sign.

KPA
Insights through analytics

Special issue: Research article

polymers advanced technologies

Received: 13 February 2015, Accepted: 22 March 2015, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/pat.3531

A multifactorial analysis of complex pharmaceutical platforms: an application of design of experiments to targetable polyacrylamide and ultrasound contrast agents

Meital Bloch[a], Ron Kenett[a]*, Lauren Jablonowski[b], Margaret Wheatley[b], Eylon Yavin[a] and Abraham Rubinstein[a]*

To improve visualization of malignant regions in the colon epithelium
cently suggested a multimodal system comprising the water-soluble c
near infrared dye derivative IR-783-S-Ph-COOH [fluorescent-cationized p
to the recognition peptide VRPMPLQ to form Flu-CPAA-Pep. The fluoresc
jugate (Flu-CPAA-Pep) is then incorporated into echogenic microbubbles
tect it from pre-mature interactions with plasma proteins upon intraveno
directed ultrasound interrogation the Flu-CPAA-Pep cargo would be rele
the MBs rupture into submicron PLA fragments (SPF). Due to their nanosc
vasculature and allow a specific binding of the Flu-CPAA-Pep to the susp

An example of finding generalization

Contents lists available at ScienceDirect

Journal of Controlled Release

journal homepage: www.elsevier.com/locate/jconrel

ELSEVIER

The effect of linker type and recognition peptide conjugation chemistry on tissue affinity and cytotoxicity of charged polyacrylamide

Meital B.D. Bloch [a], Eylon Yavin [a], Aviram Nissan [b], Ilana Ariel [c], Ron Kenett [a,d], Dovrat Brass [e], Abraham Rubinstein [a,*]

The medical problem

Colorectal cancer (CRC):

- The 3rd most common cancer diagnosed in USA.
- The 2nd leading cause of cancer-related death.

CRC treatment:

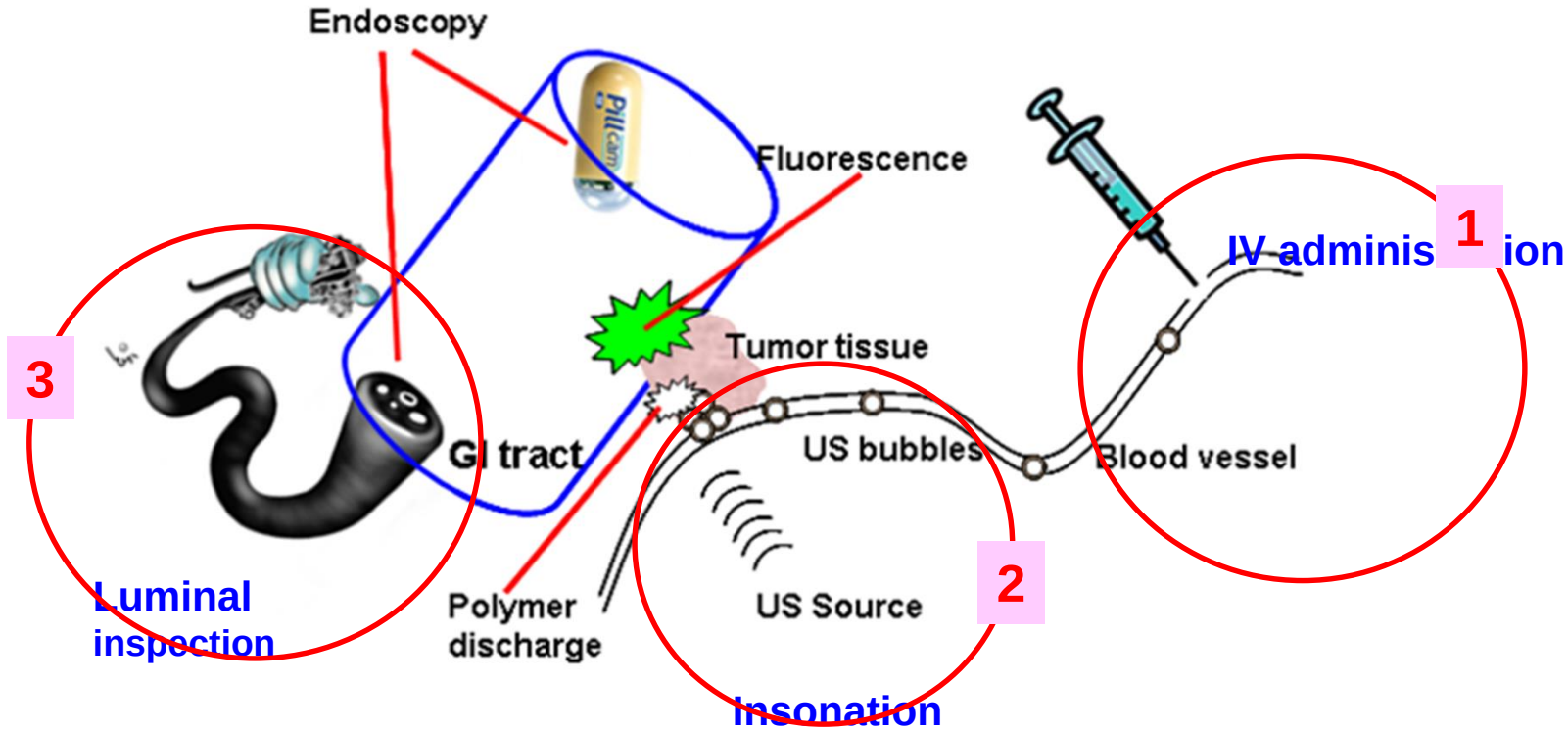
- Surgery
- Chemo/radio adjuvant therapy – depending on the CRC stage

- Overall incidence of CRC decline due to an advance in:
  - early diagnosis
  - improved medical treatments.
- This decline could even accelerate if efficient screening system is available.

Rex, D.K., *et al*. *Gastroenterology*, 112: 24, 1997.
Levin, B., *et al. Gastroenterology*, 134: 1570, 2008.
Mayer R.J. *et al*. *N. Engl. J Med,* 352: 476, 2005.
Vogelstein B. *et al*. *N. Engl. J Med*, 319: 525, 1998.
Edwards BK. *et al. Cancer, 116*: 544, 2010.

KPA Insights through analytics

The suggested concept



Hypotheses:

1. Targetability of Flu-CPAA towards dysplastic colon tissues is improved by adding a recognition peptide (Flu-CPAA-Pep).
2. Microbubbles protect Flu-CPAA and Flu-CPAA-Pep from premature affinity in the blood stream.

CPAA=cationic poly acrylamide

Power of the *in vitro* studies

Power Analysis

Significance Level	0.05
Anticipated RMSE	1

Term	Anticipated Coefficient	Power
Intercept	1	1
Mol% cat	1	1
Peptide	1	1
Presenting platform 1	1	0.988
Presenting platform 2	-1	0.917
Metastatic stage	1	0.993
Mol% cat*Peptide	1	1
Mol% cat*Presenting platform 1	-1	0.988
Mol% cat*Presenting platform 2	1	0.917
Mol% cat*Metastatic stage	-1	0.993
Peptide*Presenting platform 1	1	0.988
Peptide*Presenting platform 2	-1	0.917
Peptide*Metastatic stage	1	0.993
Presenting platform*Metastatic stage 1	-1	0.899
Presenting platform*Metastatic stage 2	1	0.84

Effect	Power
Presenting platform	0.974
Mol% cat*Presenting platform	0.974
Peptide*Presenting platform	0.974
Presenting platform*Metastatic stage	0.883

Power of the *in vivo* studies

Design Evaluation

Power Analysis

Significance Level	0.05
Anticipated RMSE	1

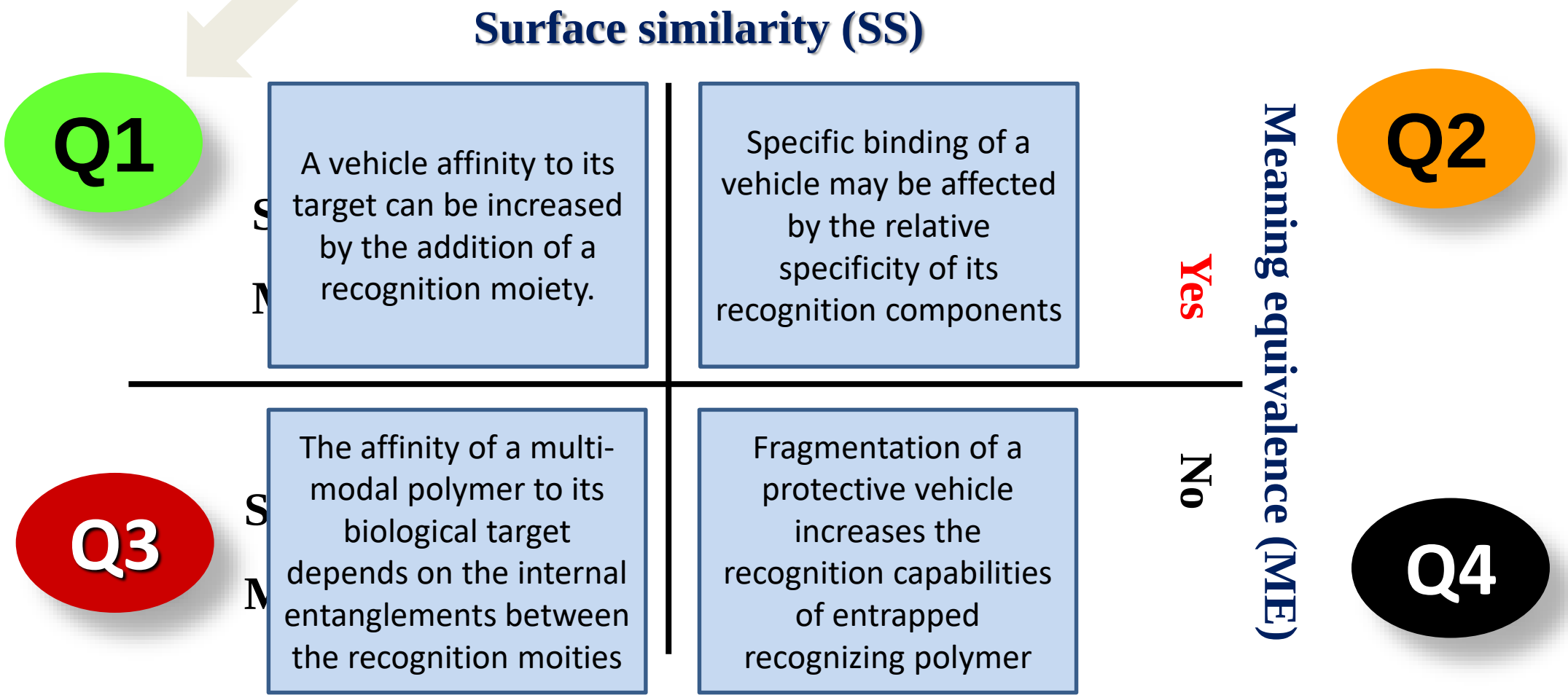
Term	Anticipated Coefficient	Power
Intercept	1	0.864
Peptide	1	0.864
Mode of administration	1	0.877
SPF	1	0.864
Peptide*Mode of administration	1	0.877
Peptide*SPF	-1	0.864
Mode of administration*SPF	1	0.877

Main Findings

1. **Increasing the charge density of Flu- CPAA-Pep leads to cross-reaction with the recognition peptide, VRPMPLQ .**
2. **Apart of Flu- CPAA-100, incorporation of the polymers into MBs did not significantly affect the MBs echogenic properties.**
3. **Flu-CPAA-Pep binds to dysplasia regions, after both IV and rectal administrations in the rat model.**
4. **Fragmenting MBs into SPF does not interfere with the affinity of Flu-CPAA and Flu-CPAA-Pep to malignant colon tissues after IV or rectal administrations in the rat.**
5. **SPF protected their Flu-CPAA-Pep cargo from non-specific interaction with serum proteins.**

Bloch M., *et al*., *Pol. Adv. Tech.* , 26: 898, 2015

Increasing the charge density leads to cross-reaction with the recognition peptide



KPA
Insights through analytics

The boundary of meaning (BOM)

Boundary of meaning

	Phrased Finding	Meaning Equivalence of the Finding (MEF) [1]	Surface Similarity Finding (SSF) [2]
1	The addition of VRPMPLQ to the Flu-CPAA backbone increased the specific binding of the polymer to their biological target.	MEF1-1: A vehicle affinity to its target can be increased by the addition of a moiety.	SSF1-1: The affinity of a multi-modal polymer to it biological tar on the internal en between the recognition moieties.
		MEF1-2: Specific binding of a vehicle may be affected by the relative spec recognition cor	SSF1-2: When one recognition moiety depends on its charge, the charge density th affinity obtained.
2	Loading the Flu-CPAA into MBs, significantly reduced the ability of the Flu-CPAA polymers to interact with their biological targets.	MEF2-1: Loading a targeted polymer into a protective vehicle interferes with the affinity properties of the polymer.	SSF2-1: Recognition polymers express reduced affinity to their biological targets when loaded into a degradable vehicle.
		MEF2-2: Recognition of a biological target by a targetable polymer depends	SSF2-2: Recognition polymer mode of loading into a protective vehicle affects the ical target.
3	Fragm SPF re properties of the Flu-CPAA polymers and even increased them.		particles ion logical ation.
		MEF3-2: Unveiling a shield from a support carrier restores the properties of the cargo polymer.	SSF3-2: Fragmentation of a protective vehicle increases the recognition capabilities of entrapped recognizing polymer.

Q1 Q3

Q1 Q3

Scientific generalization

KPA Insights through analytics