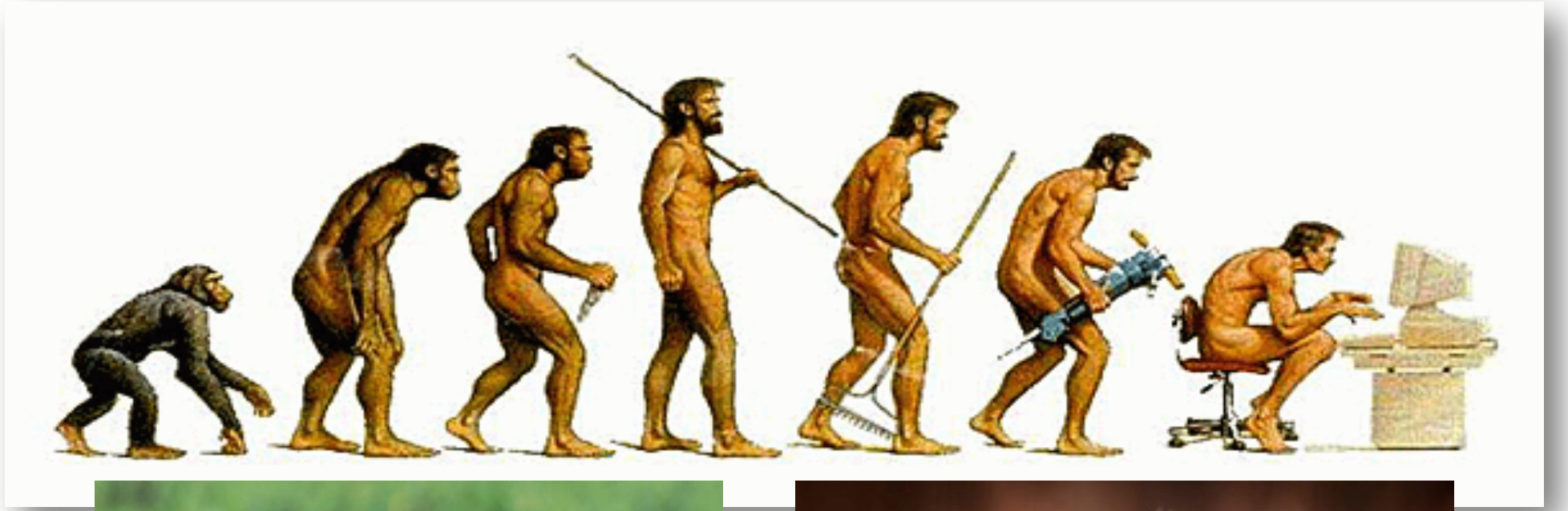
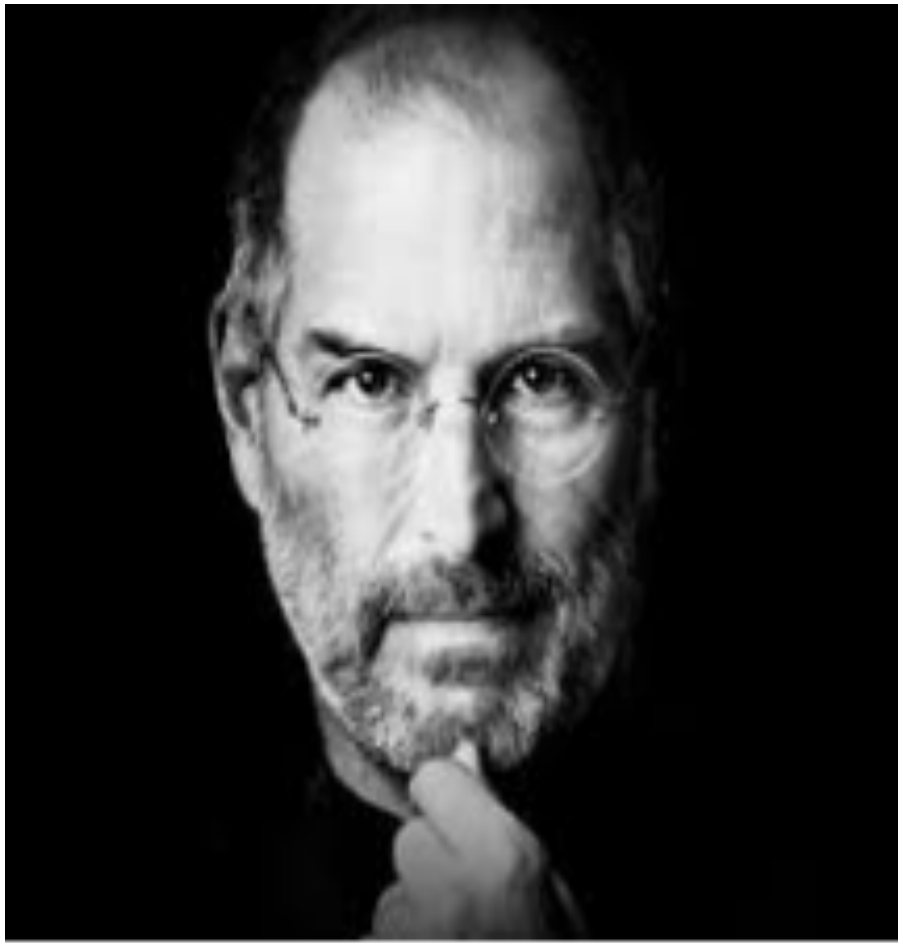
AI-Driven A/B Testing - The Next Evolutionary Leap

Steve Corney
Vaqar Khamisani
James Rubinstein

Evolution of LN Testing



Beware of HiPPOs and RATS!



AN ACTUAL HIPPO QUOTE:

"My wife was on the website last night and found the navigation confusing. You should fix that"

A former CEO* (yes, really)

not at LexisNexis!

Step 1: Remove HiPPOs and RATS!

Step 2: Get the Right KPIs and Process in Place

Example: Schipol Airport, Amsterdam

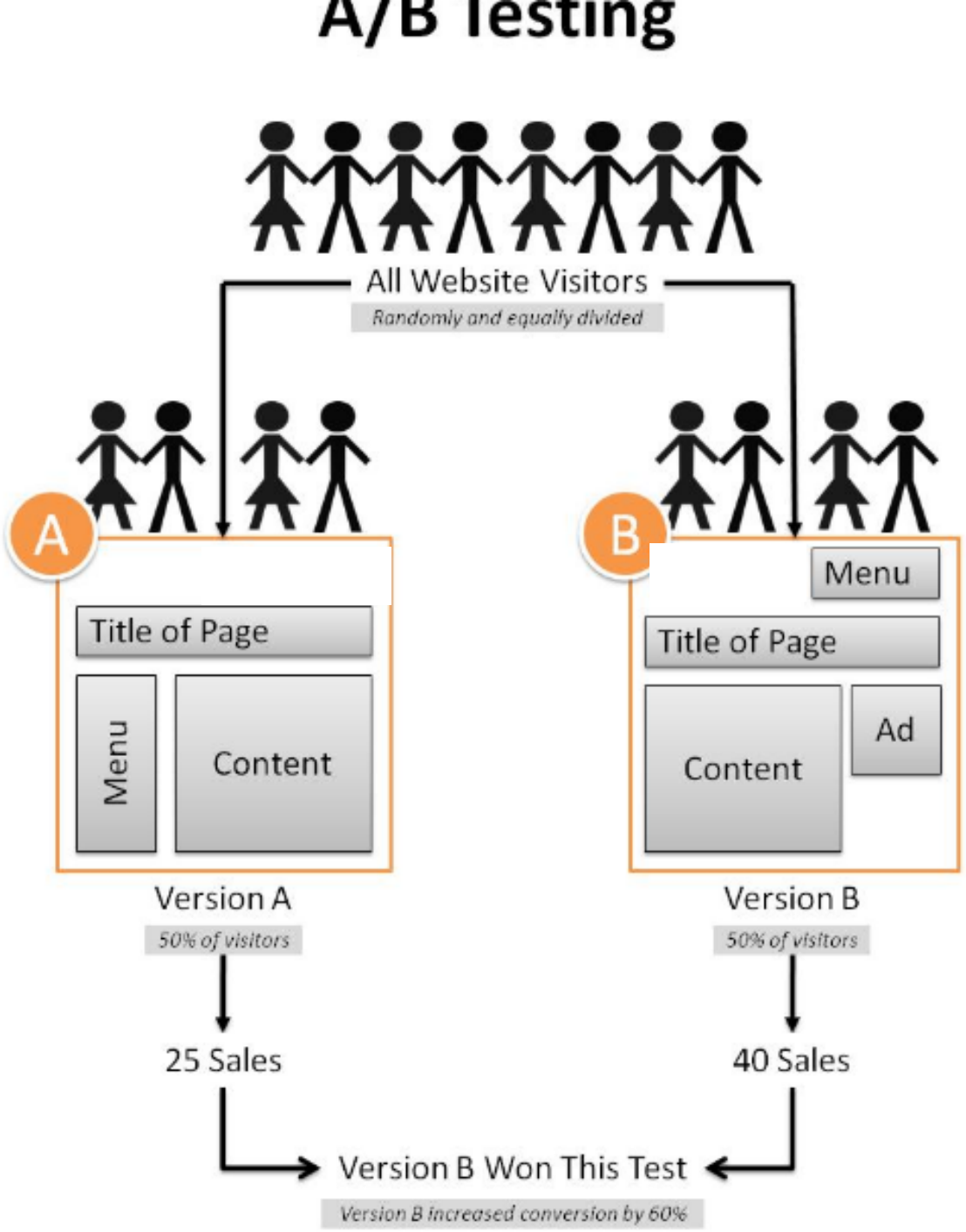
- **Define a goal** (lower cost of cleaning)
- **Find the one KPI to measure** (spillage rate)
- **Define a test hypothesis** to change this KPI
- **Test:** Designers observed an inherent male behavioural instinct to 'aim' and strategically placed a small printed 'fly' to the urinal

Result:

- 85% less spillage
- €3m a year saved in cleaning bills

But even when the culture and processes are right…

AI-Driven AB Testing
Approach and Methodology



A/B Testing Key Issues

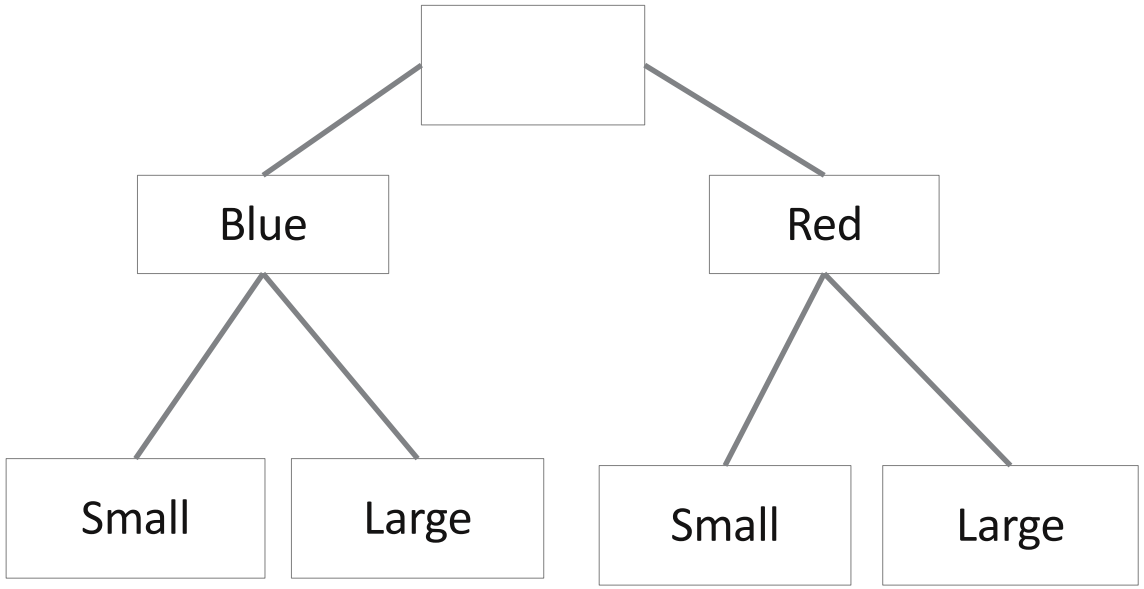
Slow Speed:

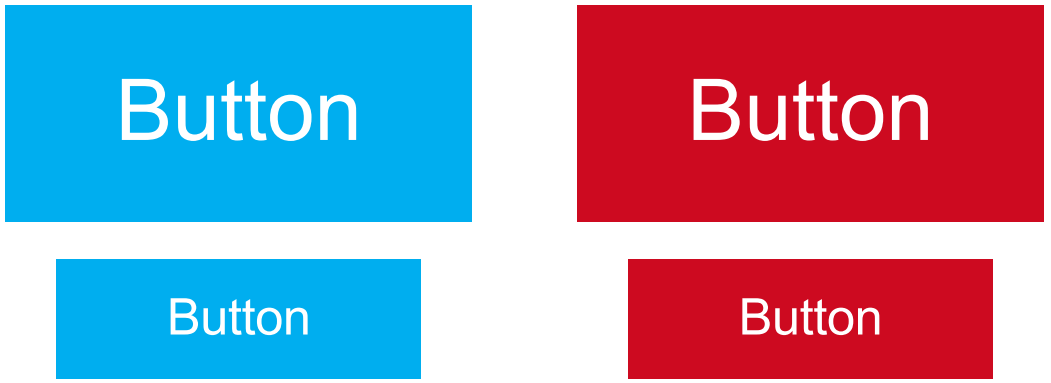
A/B testing with one/few improvements at a time is a slow process. As the number of changes are small in each iteration, several such cycles need to happen before any meaningful impact.

Testing to Local Optimum: Running a testing program with a single or a few changes in each test can lead to sub-optimal results over time

Local vs. global optimum: button colour and size

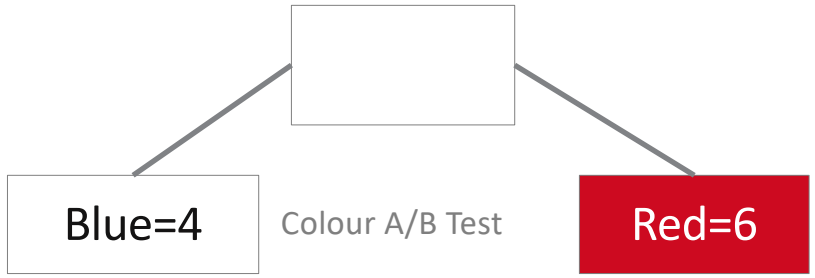
Which is the best combination of blue, red, large and small?





Local vs. global optimum: button colour and size

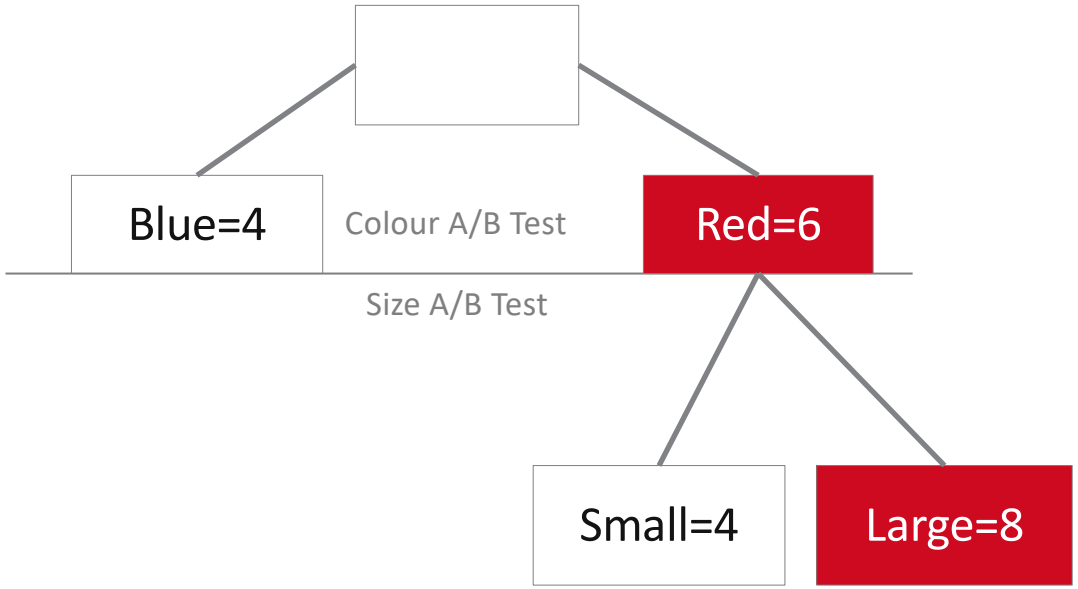
Which is the best combination of blue, red, large and small?



Testing colour only returns test winner = RED

Local vs. global optimum: button colour and size

Which is the best combination of blue, red, large and small?



Testing colour only returns test winner = RED

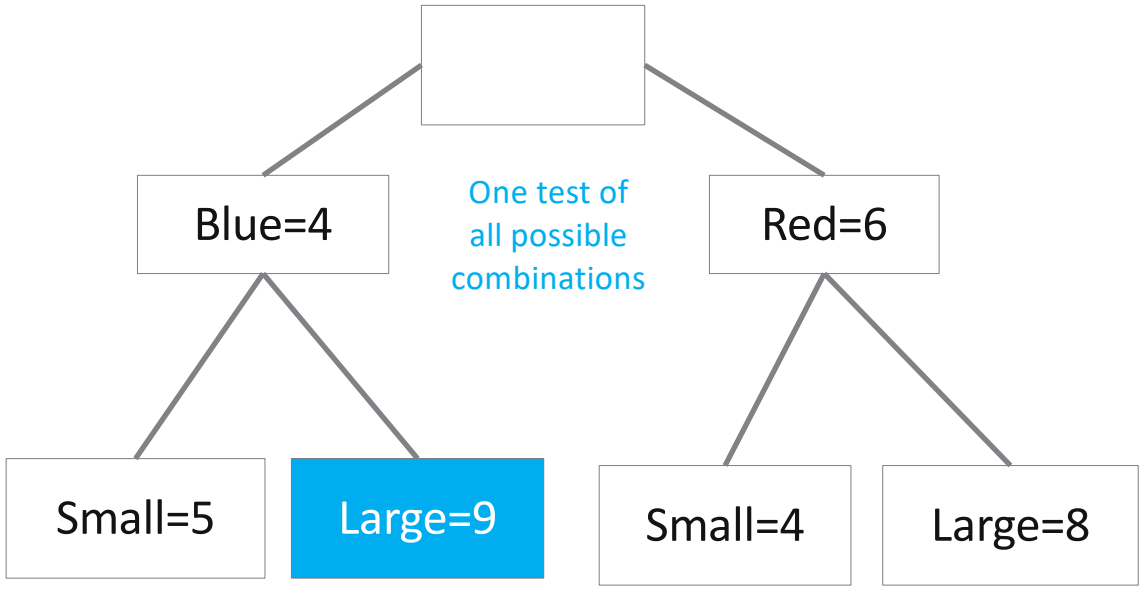
The next test for size returns test winner = LARGE

Sequential a/b testing (first colour, then size) would result in RED the winner of test 1 and then LARGE the winner of test 2.

KPI = 8

Local vs. global optimum: button colour and size

Which is the best combination of blue, red, large and small?



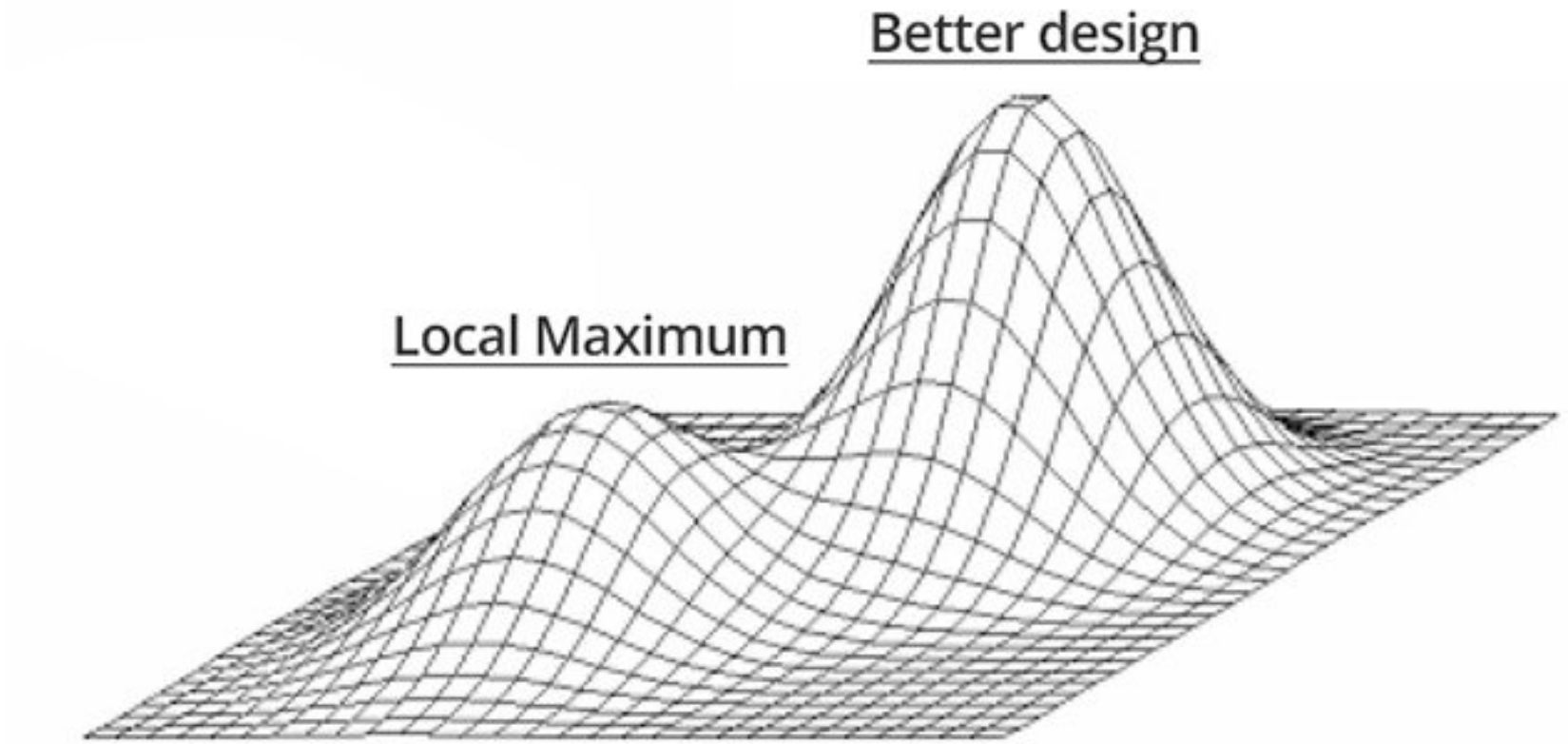
However – by testing in small increments, we identify the local optimum and miss out on the opportunity for a potentially bigger global optimum

A single multivariate test which evaluates all variations at the same time (red small, red large, blue small, blue large) would return a winner of

BLUE + LARGE

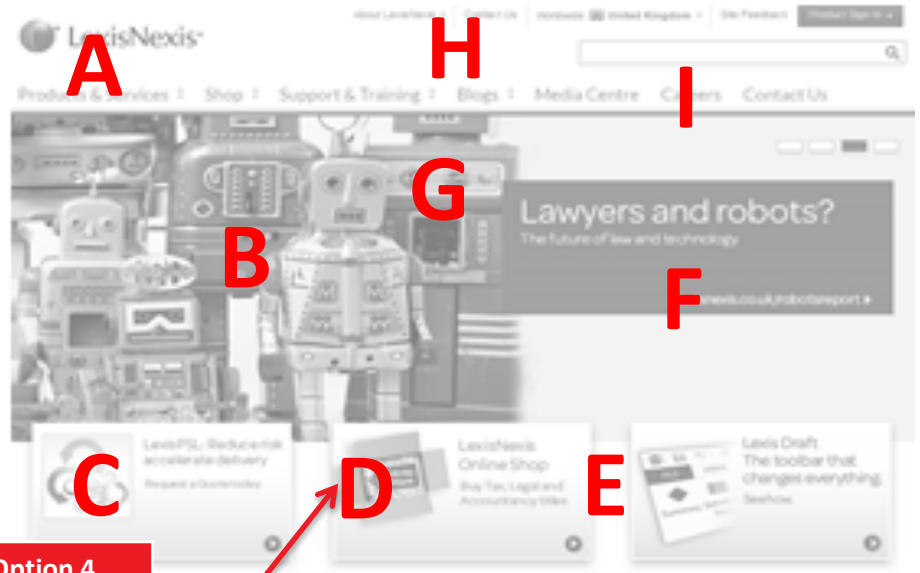
KPI = 9

Multivariate testing will achieve global optimum



Challenges of Multivariate Testing

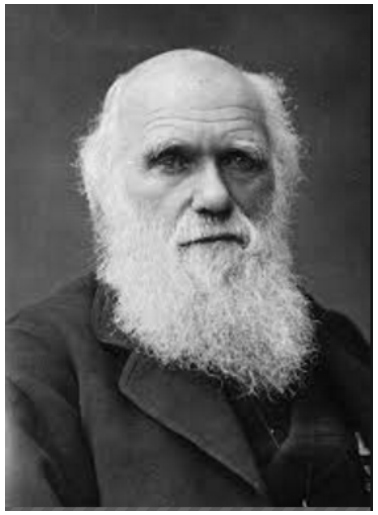
- Multivariate testing can achieve the needed acceleration as well as optimality.
- However, large numbers of combinations and traffic limitations make it impractical for product testing.



Area	Option 1	Option 2	Option 3	Option 4
A	A1	A2	A3	A4
B	B1	B2	B3	B4
C	C1	C2	C3	C4
D	D1	D2	D3	D4
E	E1	E2	E3	E4
F	F1	F2	F3	F4
G	G1	G2	G3	G4
H	H1	H2	H3	H4
I	I1	I2	I3	I4

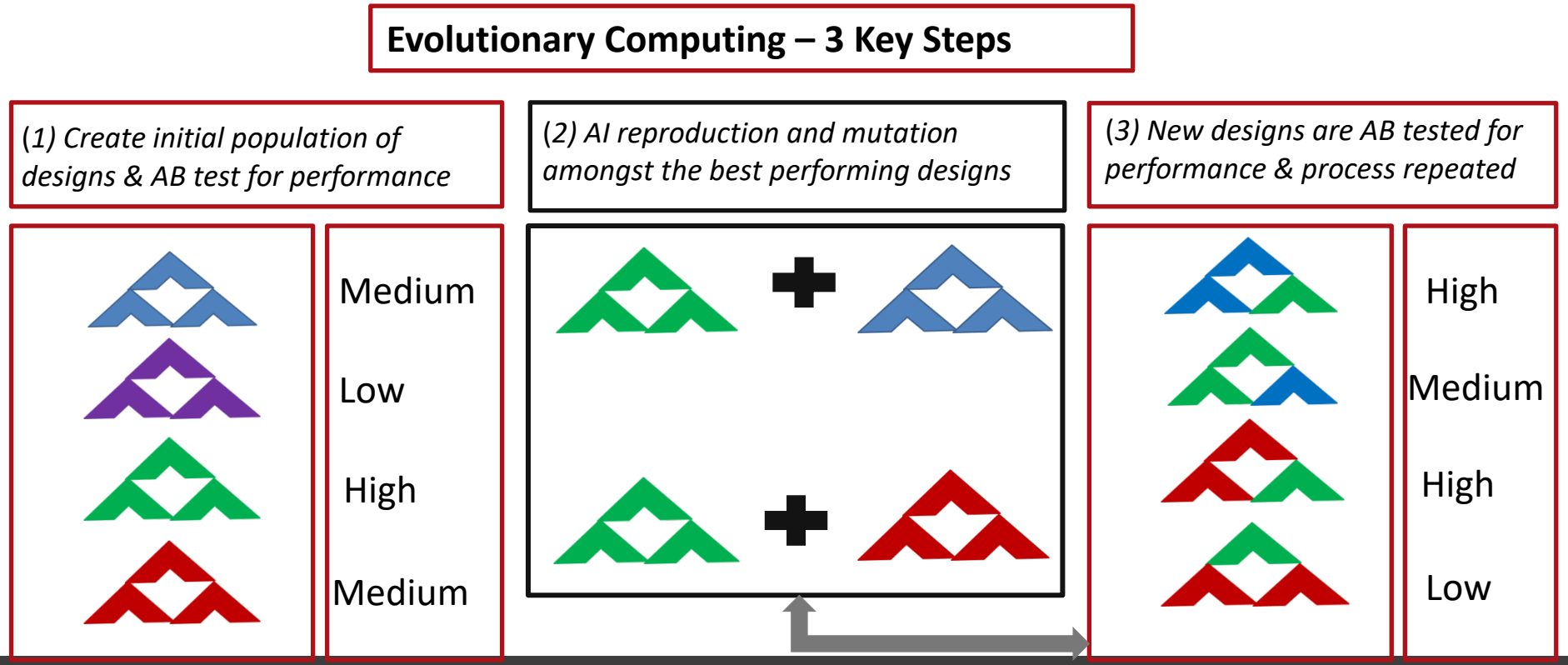
Total Combinations 262,144

The Solution?



Evolutionary Computing

- Evolutionary computing can be designed to test only a fraction of the whole design space and provide a near optimal solution.
- This solution addresses the issues of scale and speed as well as optimality through machine learning



US Store Pilot

Executive Test Summary

Goal

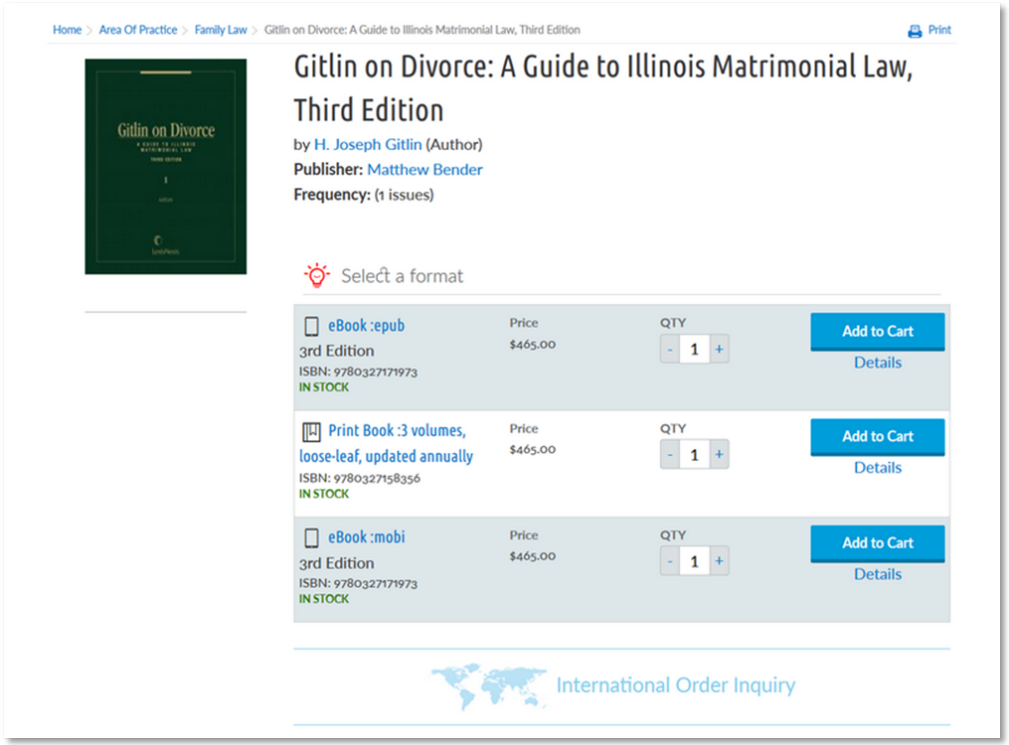
- Evaluate the application of machine learning on MVT
- Increase “Add to cart” clicks on the product page
- Increase the clicks to product details

Hypothesis

- Evolutionary testing will rapidly produce an optimised design which improves KPIs

Key Performance Indicators (KPIs)

- % Add to cart clicks (60%)
- % Clicks to product details page (40%)



Population and gene pool

- A/B Testing tool created all possible element variations (22)
- Evolutionary algorithm determines each generation
- LexisNexis configured genomes to be tested in each wave based on EA output.
- Blue genome is the control (default)

Number of possible combinations: 184,320

A	B	C	D	E	F	G	H	I	J	K	L	M
Add to Cart Colour	Add to Cart Label	Add to Cart Icon Design	Detail label	Details Link Style	Format Link Design	International Order Label	Sample Chapter Txt	Sample Chapter Icon	Sample Chapter Location	ToC Design	Print Icon	Trustwave Icon
Blue	Add to Cart	No Icon	Details	Text Link	Zebra Stripes	International Order	Long Text	No Icon	Within Product	Two Columns	Present	Footer
Red	Buy Now	Arrow Square	View Details	Button	No Zebra Stripes	Outside the US	Simple Text	Glasses	Above Thumbnail	One column	Remove	Under Thumbnail
Green	Order Now	Arrow Circle	More Information					Forward				
Orange		Plus Square						Book Stack				
		Plus Circle										

Wave (generation) summary

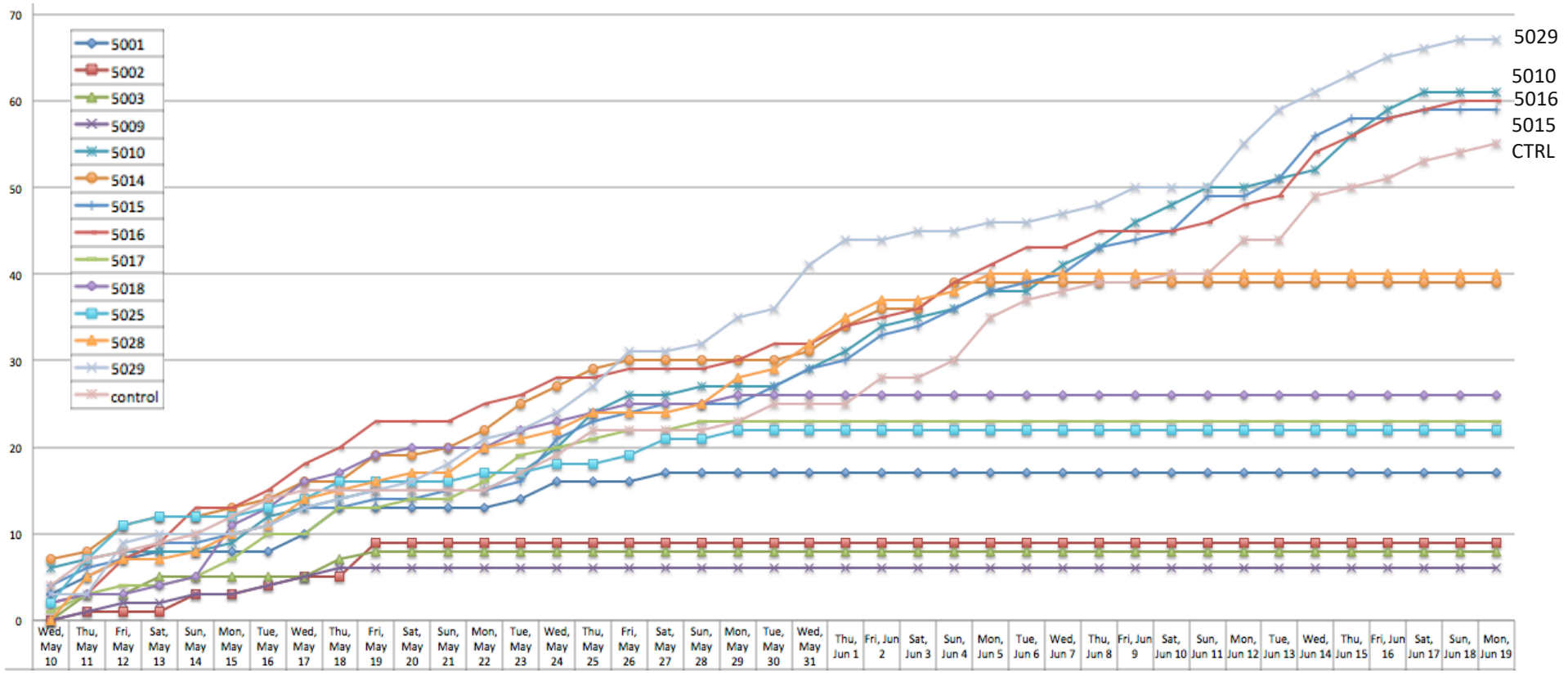
Wave / generation	Genomes	Views	Visitors	Add to Cart (UAR%)	View Detail (UAR%)	Duration
1 (28 March – 04 April)	50	4,946	3,203	90 (2.81%)	143 (4.46%)	7 days
2 (04 April – 07 April)	30	5,909	3,608	107 (2.97%)	197 (5.46%)	3 days
3 (07 April – 12 April)	30	6,877	4,214	109 (2.59%)	181 (4.30%)	5 days
4 (12 April – 19 April)	27	9,584	5,921	170 (2.87%)	265 (4.48%)	7 days
5 (19 April – 10 May)	25	30,073	18,687	387 (2.07%)	788 (4.22%)	21 days
6* (10 May – 19 June)	14	41,757	24,746	490 (1.98%)	1,250 (5.05%)	40 days

TOTALS:	Genomes	Views	Visitors	Cart Adds	Detail Views	Duration
	162	99,146	60,379	1,353	2,824	83 days

* Wave 6 was a validation wave of wave 5. The top 13 performing creative from wave 5 plus control were started as a new wave. This did not introduce any new variations, but further evaluated the existing generation from 5.

While Wave 6 was running, the lowest performers were dropped every 4 days.

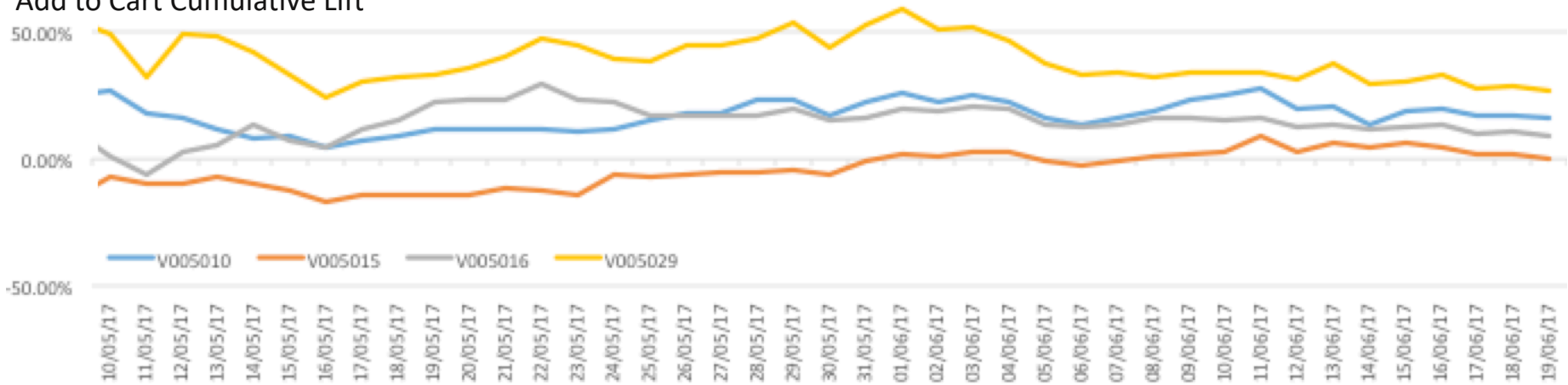
Wave 6: Cumulative daily clicks



Wave 6: Details

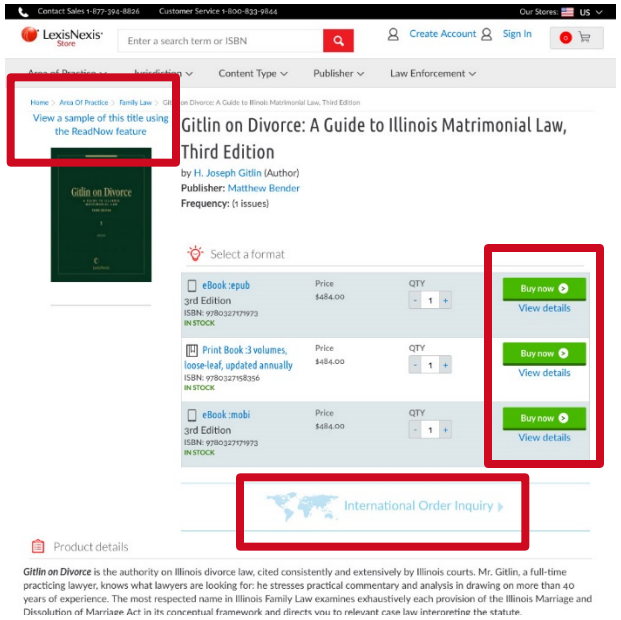
Challenger	Views	Visitors	Add to Cart (UAR%)	lift	View Detail (UAR%)	lift	Combined UAR%	Combined lift
control	5,268	3,123	55 (1.76%)	-	145 (4.64%)	-	2.90%	-
5029	5,343	3,151	66 (2.09%)	18.93%	182 (5.78%)	24.40%	3.58%	23.23%
5016	5,411	3,088	59 (1.91%)	8.49%	176 (5.70%)	22.76%	3.43%	18.13%
5015	5,134	3,055	58 (1.90%)	7.80%	160 (5.24%)	12.80%	3.25%	11.99%
5010	5,346	3,161	61 (1.93%)	9.58%	159 (5.03%)	8.34%	3.18%	9.72%

Add to Cart Cumulative Lift

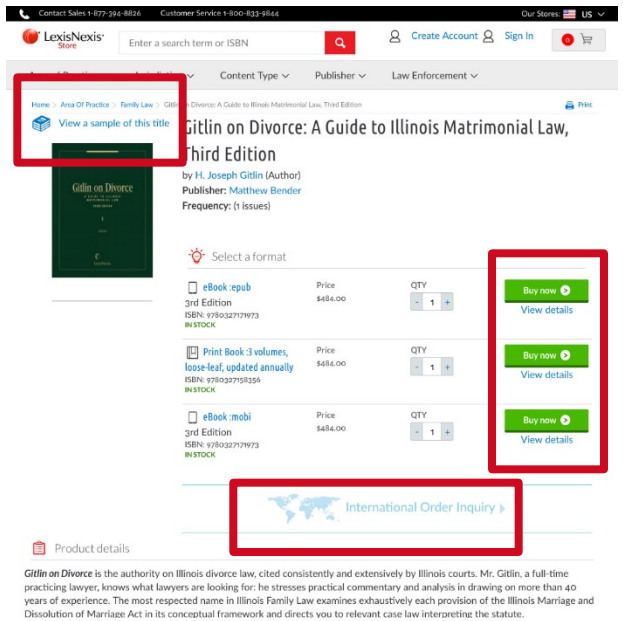


Wave 6: Top performer details

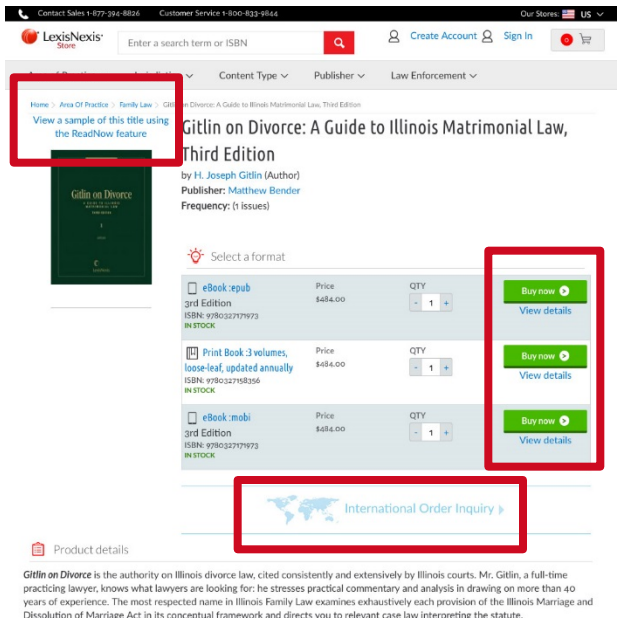
5029



5010



5016



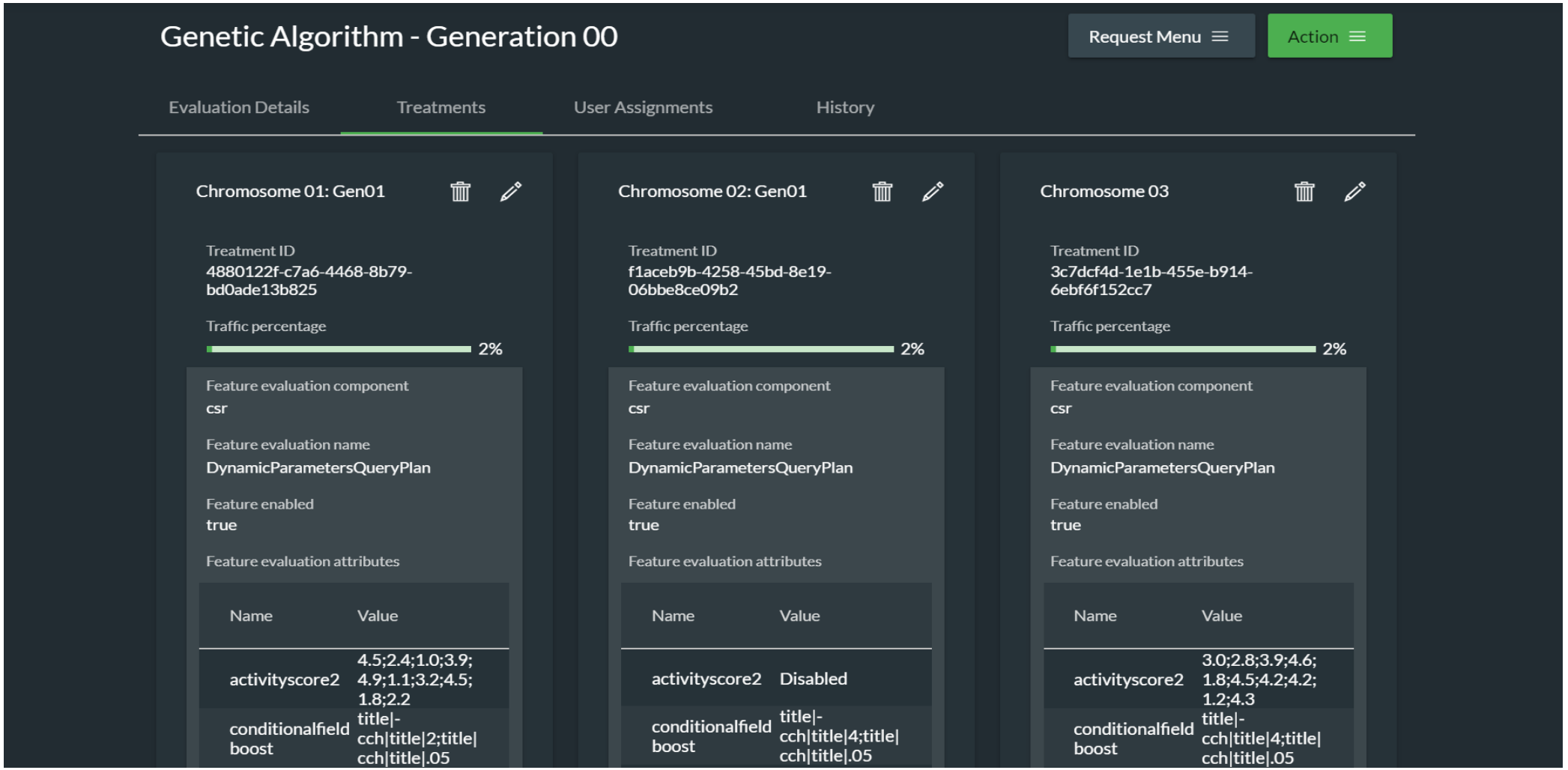
Genome	cart colour	cart label	cart icon	details link	details style	format bg	intl order	sample copy	sample icon	sample loc	Toc style	print icon	trust loc
V005029	green	buy now	arrow Circle	view details	c	c	c	c	c	above thumb	c	hide	c
V005016	green	buy now	arrow Circle	view details	c	c	c	c	c	above thumb	one Col	hide	c
V005010	green	buy now	arrow Circle	view details	c	no stripes	c	simple	book Stack	above thumb	c	c	c

8 of the 13 elements are identical in the top performers

The top two performers are virtual twins with only one differing element


20%

What's Next


Genetic Algorithm - Generation 00
Request Menu
Action
Evaluation Details
Treatments
User Assignments
History
Chromosome 01: Gen01
Treatment ID
4880122f-c7a6-4468-8b79-bd0ade13b825
Traffic percentage
2%
Feature evaluation component
csr
Feature evaluation name
DynamicParametersQueryPlan
Feature enabled
true
Feature evaluation attributes
Name
Value
activityscore2
4.5;2.4;1.0;3.9;4.9;1.1;3.2;4.5;1.8;2.2
conditionalfield boost
title|-cch|title|2;title|cch|title|.05
Chromosome 02: Gen01
Treatment ID
f1aceb9b-4258-45bd-8e19-06bbe8ce09b2
Traffic percentage
2%
Feature evaluation component
csr
Feature evaluation name
DynamicParametersQueryPlan
Feature enabled
true
Feature evaluation attributes
Name
Value
activityscore2
Disabled
conditionalfield boost
title|-cch|title|4;title|cch|title|.05
Chromosome 03
Treatment ID
3c7dcf4d-1e1b-455e-b914-6ebf6f152cc7
Traffic percentage
2%
Feature evaluation component
csr
Feature evaluation name
DynamicParametersQueryPlan
Feature enabled
true
Feature evaluation attributes
Name
Value
activityscore2
3.0;2.8;3.9;4.6;1.8;4.5;4.2;4.2;1.2;4.3
conditionalfield boost
title|-cch|title|4;title|cch|title|.05

We’re Hiring!

Data Scientists

Data Engineers

SWEs

James.Rubinstein@lexisnexis.com

Questions