How To Run Many Tests At Once

Interaction Avoidance and Detection

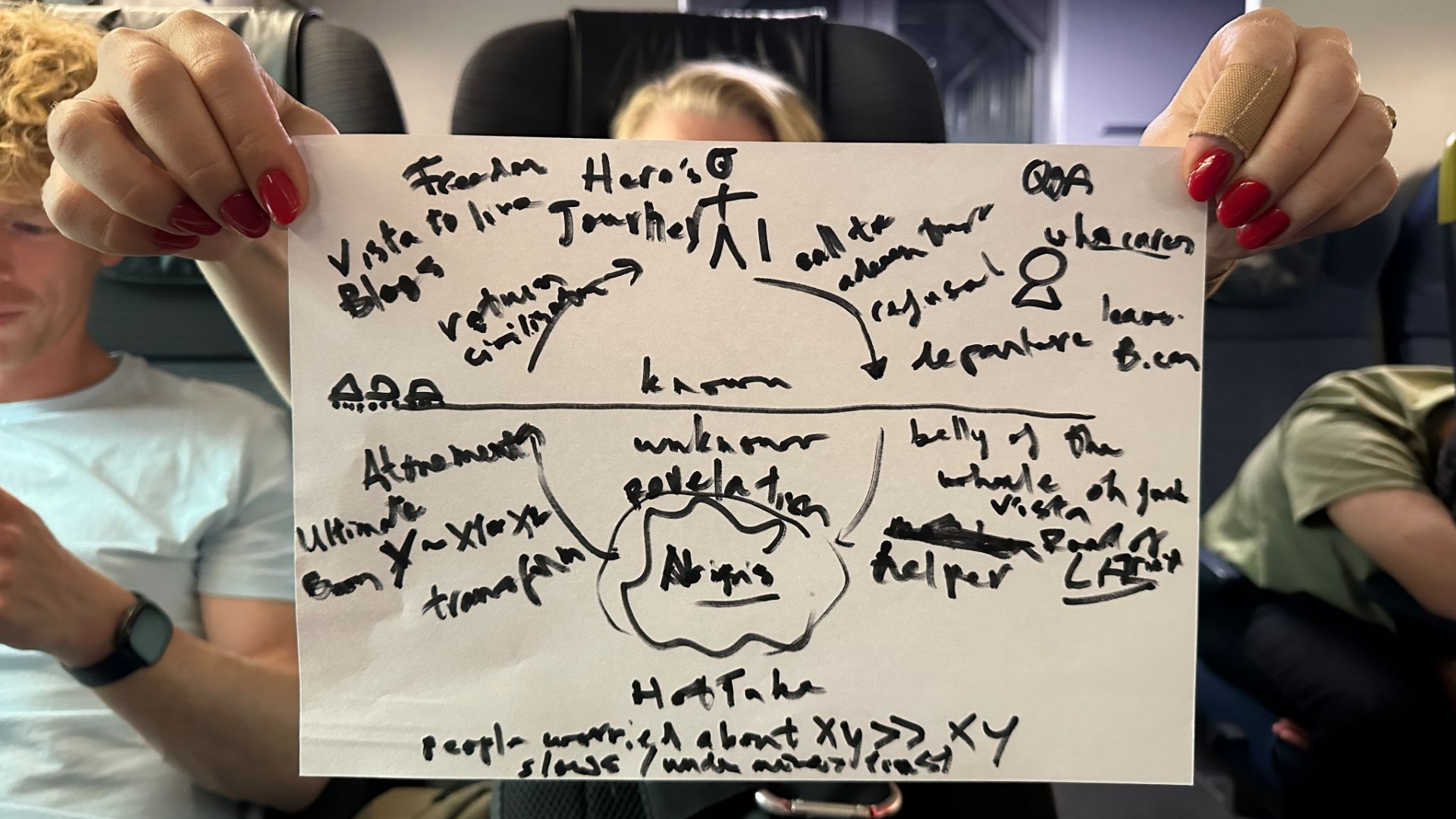
Lukas Vermeer
DDMA Experimentation Heroes, October 31st, 2023

Alexander Richter (He/Him) • 2nd

ABlyft - A/B-Testing Platform

1d •••

That is specifically one of the talks I'm really looking forward to **Lukas Vermeer**


Freedom
Hero's
Journey
Vista to live
Blogs
call to adventure
refusal
departure
known
unknown
revelation
belly of the whale
oh fuck
helper
Origin
Ultimate
transform
HotTake
people worried about XY >> XY

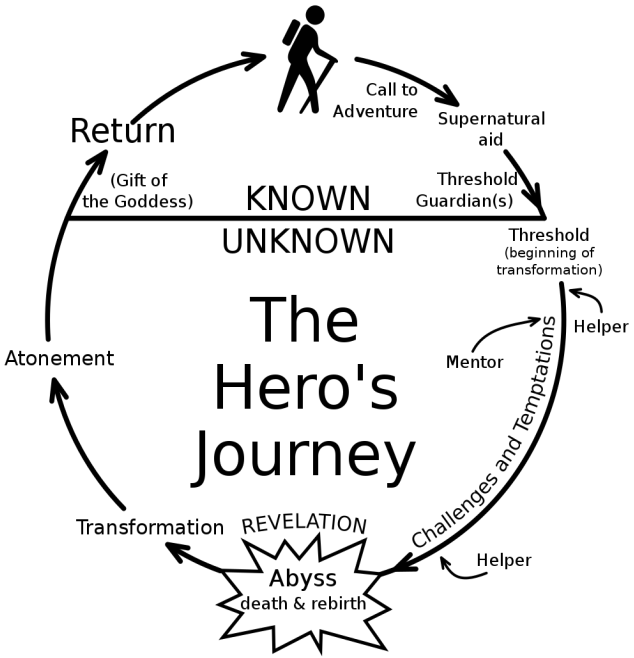
Something Something Experimentation Something

The monomyth of pretty much every talk I've ever done.

Lukas Vermeer
Some event, Somewhere

A personal story

A first person narrative involving failure and learning to connect with the audience.



Some

Mind

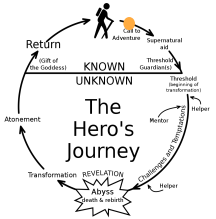
Blowing

Content

Obligatory Q&A

Your opportunity to ask me pretty much anything at all.

“But what about the interactions!?”

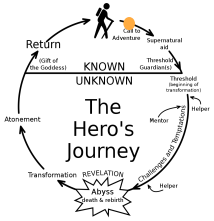


1. The Call to Adventure

Yes. What about them?

As long as experiments are executed orthogonally (fancy word alert!) and there are no interaction effects, The Math Just Works™.

	Users in A of experiment 1	**Users in B of experiment 1**
Users in A of experiment 2	control	+5%
Users in B of experiment 2	+10%	+15%

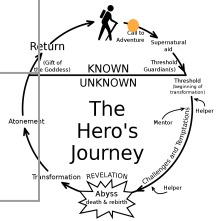


1. The Call to Adventure

Yes. What about them?

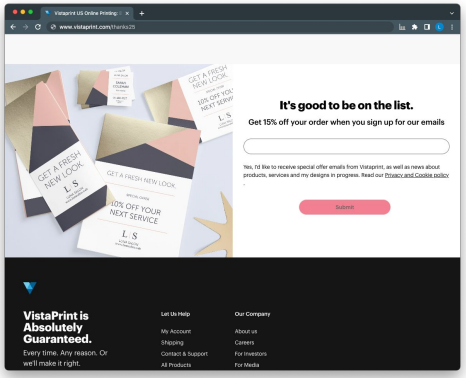
As long as experiments are executed orthogonally (fancy word alert!) and there are no interaction effects, The Math Just Works™.

	Users in A of experiment 1	**Users in B of experiment 1**	**What we see experiment 2**
Users in A of experiment 2	control	+5%	control (0 + 5) / 2 = 2.5%
Users in B of experiment 2	+10%	+15%	+10% (10 + 15) / 2 = 12.5%
What we see experiment 1	control (0 + 10) / 2 = 5%	+5% (5 + 15) / 2 = 10%	

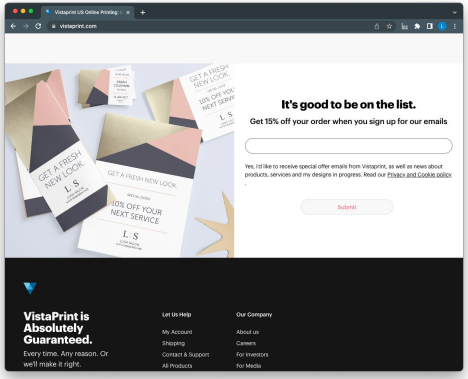


1. The Call to Adventure

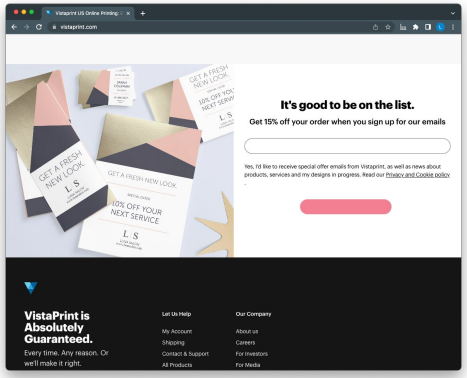
A silly example to illustrate (functional) interactions

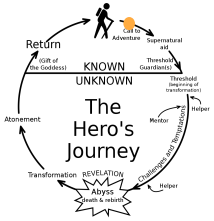


+



=

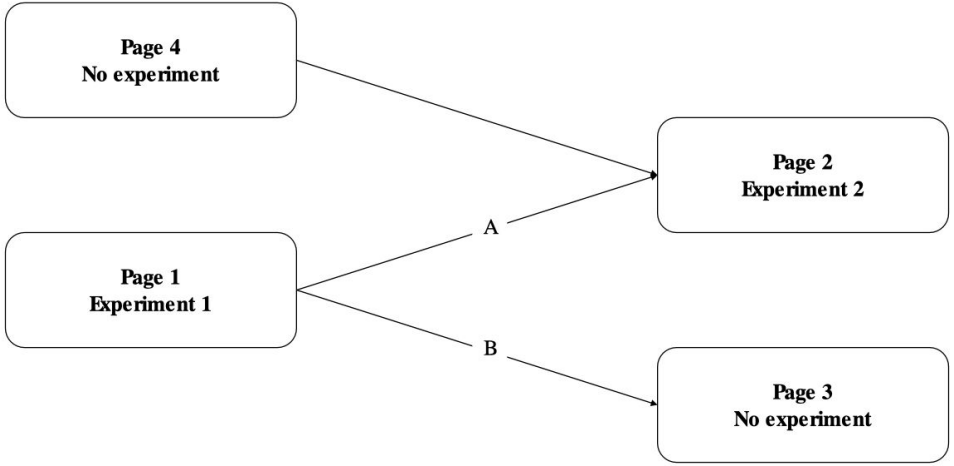


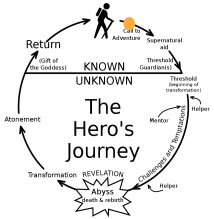


1. The Call to Adventure

Two kinds of interactions: traffic interactions

One experiment treatment causes a different mix of traffic to flow to another experiment. A statistician would call this “sampling bias”.



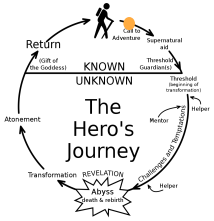


1. The Call to Adventure

Two kinds of interactions: metric interactions

Impact on a metric for a combination of two experiments differs from what we see in either experiment in isolation.

	Users in A of Experiment 1	Users in B of Experiment 1
Users in A of Experiment 2	Control	+5%
Users in B of Experiment 2	+10%	-50%



1. The Call to Adventure

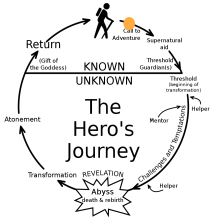
Potential consequences

1. Biased measurement

Bias Example 2	Effect
A of Experiment 2 (50% of users in experiment 1)	+10%
B of Experiment 2 (50% of users in experiment 1)	+20% —B doubles the effect of experiment 1!
Total for Users in Experiment 1	+15%

2. Inference and decisions errors

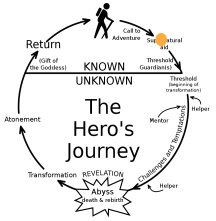
Bias Example 1	Effect
A of Experiment 2 (50% of users in experiment 1)	+10%
B of Experiment 2 (50% of users in experiment 1)	-100%
Total for Users in Experiment 1	-45%



1. The Call to Adventure

Interaction effects are not a problem worth worrying about

They are rare[~~citation needed~~ 1] and severe ones are obvious and easy to avoid or detect.



2. Refusal of the Call

lukasvermeer.medium.com

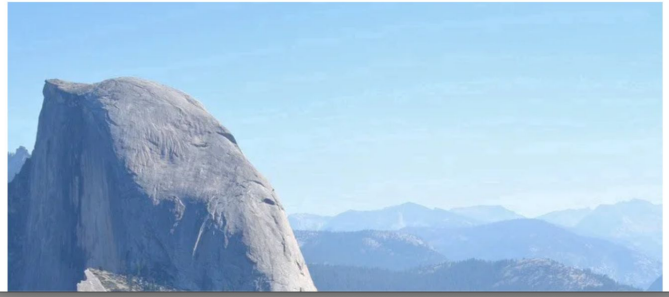
Why I am leaving a place I love

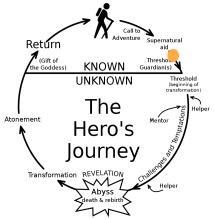
FOMO made me do it

Lukas Vermeer · Follow
3 min read · Apr 13, 2021

At the end of this month, I will be leaving the company where I have spent the past eight amazing years. I am not sure I can adequately explain my motivation (decisions of this nature rarely have a single cause) but I will try.

Not all decisions in life can be data driven. I am leaving because it simply felt like it was time for me to leave. I will use three metaphors to try to convey how I feel. I hope this may help others who are in a similar position.





4. The Crossing of the First Threshold

lukasvermeer.medium.com

Choosing my own adventure

Three things I wanted to experience in my next job

Lukas Vermeer · Follow
4 min read · Jun 1, 2021

60

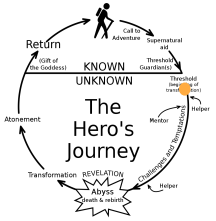
One month ago, I started a new job. I left a place I love to go exploring. After eight years of following a clear path, I was ready to be left to my own devices and seek out a new course.

What came after was to choose the next step. I did not want to leave one place without going to another destination. Deciding where to go next was almost as difficult as deciding to leave. But I knew that—more than anything specific—I wanted something different, something new, and some new things.

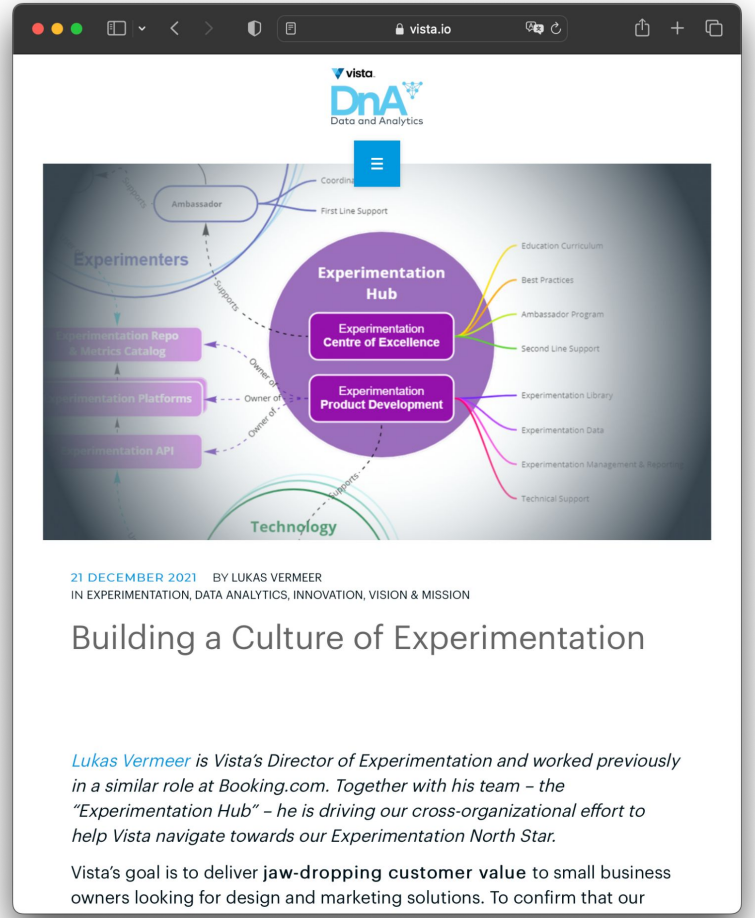
A new challenge. A new perspective. A new setting.

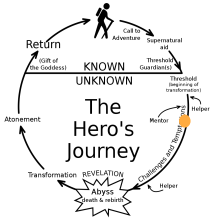
I understand and am grateful for the fact that I am speaking from a place of immense privilege. I had options to choose from. This is not the case for many others. Still, I hope this post may help others who have the luxurious challenge of having to choose their own adventure.

I chose Vistaprint, and here's why



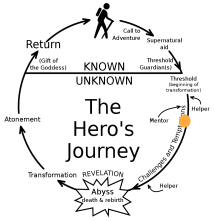
5. Belly of the Whale





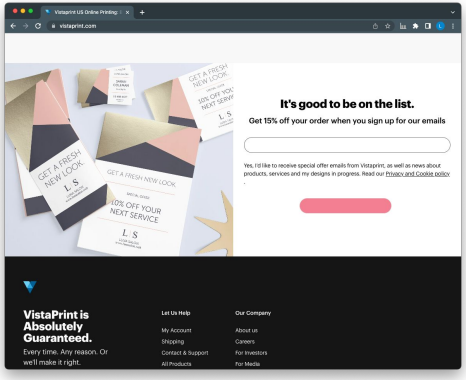
6. The Road of Trials

“But what about the interactions!?”



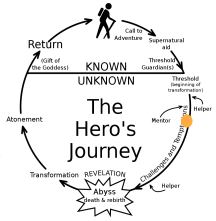
6. The Road of Trials

Null Strategy: Not Running an Experiment at All



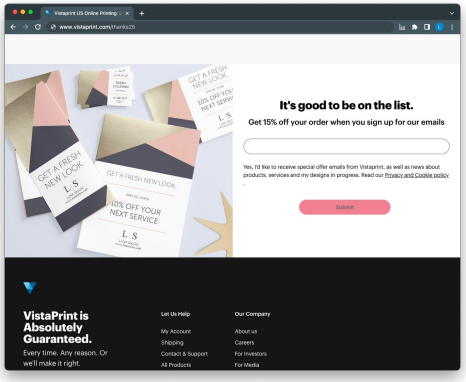
“Just ship it!”

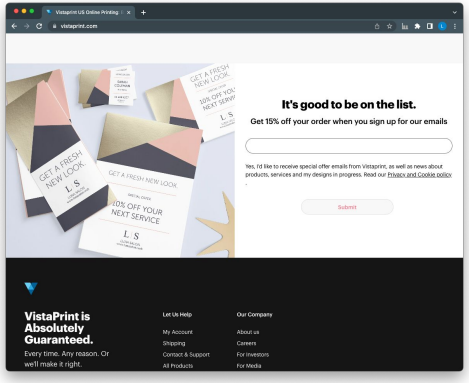
time



6. The Road of Trials

Sequential Avoidance: Running One Experiment at a Time





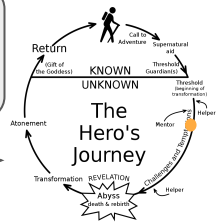
"First try this."

"Then try this."

Experiment 1

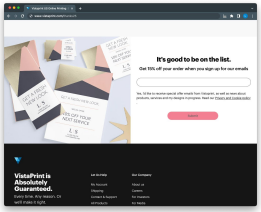
Experiment 2

time



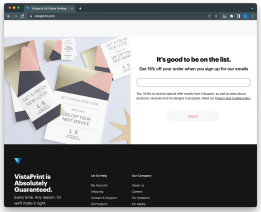
6. The Road of Trials

Isolation Avoidance: Running Experiments in Separated Lanes



"We try this."

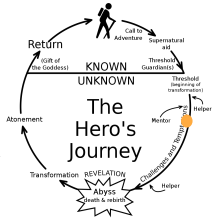
Experiment 1



"They try this."

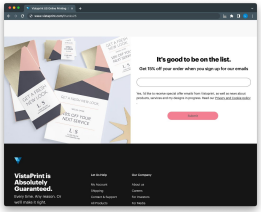
Experiment 2

time

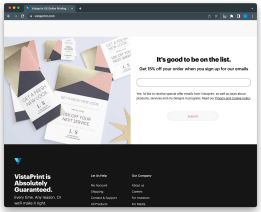


6. The Road of Trials

Combined Avoidance: Combining Treatments Into a Single Experiment



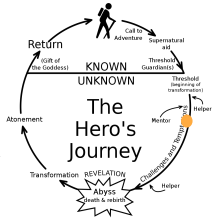
"This is B."



"This is C."

ABC Experiment

time

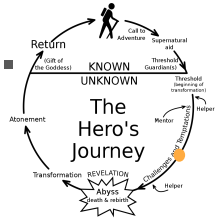


6. The Road of Trials

People being worried about interaction effects is a problem worth worrying about

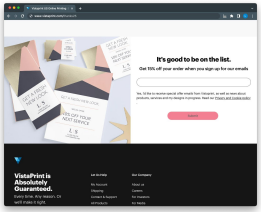
Worrying undermines trust and reduces velocity.

Dogmatic avoidance has a very high cost.

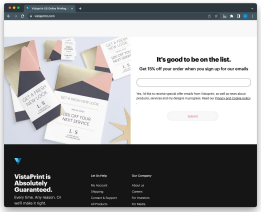


7. The Meeting with the Goddess

Detection Instead: Not Avoiding Interactions and Choosing Detection



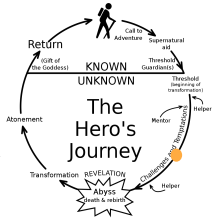
“We try this.”



“They try this.”

Experiment 1 and also Experiment 2
(+interaction detection)

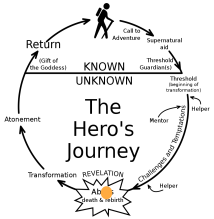
time



7. The Meeting with the Goddess

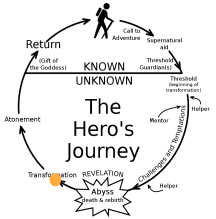
Uh…

And then I realised I had no idea what I was actually talking about all those years.



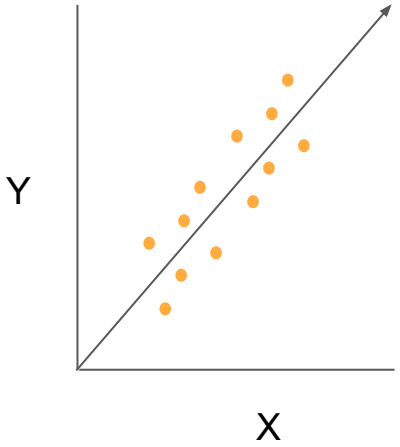
9. Abyss

“Why don’t you just use a regression?”



10. Apotheosis

“Why don’t you just ~~use a regression~~ draw a line?”



Regression tries to fit a line to some data given a model.

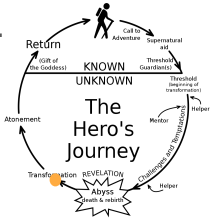
In this example, the model given is Y ~ X. In other words, estimate Y given only X.

This model describes a linear function of the form Y = a + X * b. The regression will try to estimate a and b.

The line fits our data best when a=0 and b=1.

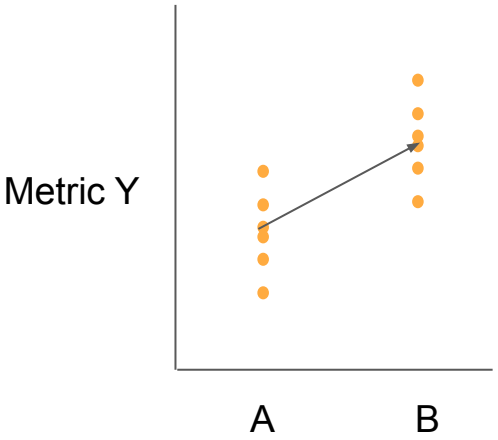
Regression works through mathemagic.

Please don’t ask me to explain.



10. Apotheosis

Aside: connecting the dots



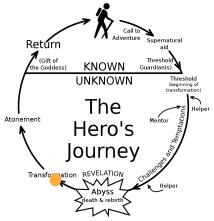
The exact same model could be used to analyse the results of a regular A/B test.

The linear function $Y = a + X * b$ could represent the results of a test if we assume

- Y = the metric of interest
- a = base rate
- X = which treatment user was exposed to
- b = the effect of the treatment

Most regression implementations will return confidence intervals and p-values for b.

You probably don't want to do this.

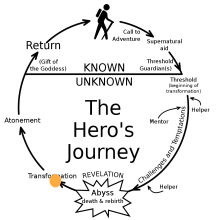


10. Apotheosis

Y ~ X1 * X2

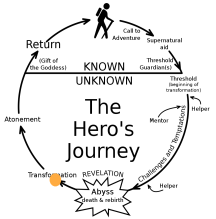
We can extend the regression model to include more than one input. We can also include “interaction terms” which combine multiple inputs.

The above model expands to Y = a + X1 * b + X2 * c + X1:X2 * d



10. Apotheosis

```
> summary(lm(metric_value ~ exp_1*exp_2, data = df))

Call:
lm(formula = metric_value ~ exp_1 * exp_2, data = df)

Residuals:
Min   1Q Median   3Q    Max
-25.38 -22.34 -20.82 -20.20 280.00

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.1959  0.3081  65.543  < 2e-16 ***
exp_1        0.6201  0.4339   1.429   0.153
exp_2        2.1495  0.4348   4.943 7.69e-07 ***
exp_1:exp_2   2.4183  0.6146   3.935 8.33e-05 ***
---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.59 on 99996 degrees of freedom
Multiple R-squared:  0.001691,   Adjusted R-squared:  0.001661

 F-statistic: 56.45 on 3 and 99996 DF,  p-value: < 2.2e-16
```



10. Apotheosis

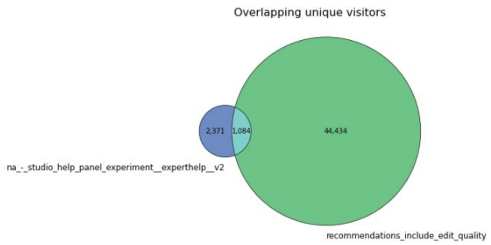
EXPO-1328 [Recommendation: | XY Interaction Detection - Dat:

File | /Users/Lukas.Vermeer/Downloads/interaction_detection_report_1414009.html

databricks

Import Notebook

Experiment Interaction Detection



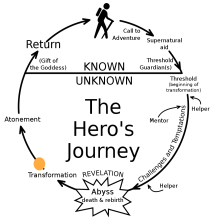
Traffic Interaction Results:

Metric Interaction Summary Results:

No significant result.
There is no evidence for an interaction effect. P-value: 0.2571

METRIC_G...	EMA_EVENT_METRIC	TARGET_METRIC	ARE_INTERACT
Primary Metric	Configure - Recommendation-Cross sell-Engagement	UNIQUE_CONVERSION_PER_VISITOR	No
Primary Metric	Configure - Recommendation-Cross sell-Engagement	TOTAL_CONVERSION_PER_VISITOR	No
Conversion	01-Order Completed-Total Revenue-USD	REVENUE	No
Conversion	01-Order Completed-Total Revenue-USD	TOTAL_CONVERSION_PER_VISITOR	No
Conversion	01-Order Completed-Total Revenue-USD	VGP	No
Conversion	01-Order Completed-Total Revenue-USD	UNIQUE_CONVERSION_PER_VISITOR	Yes
Conversion	01-Order Completed-Total Revenue-USD	VTP	No

Metric Interaction Results:



11. The Ultimate Boon

EXPO-1328 [Recommendation

staging.vistaprint.com/experimentation/experiments/EXPO-1328

vista Experimentation Hub

MANAGEMENT

- Experiments
- Feature Flags
- Metrics
- QA

Get Help

You Think You Can Test?

LV Lukas Vermeer

EXPO-1328 [Recommendations] Improve matching by selecting it depending on the edit quality score instead of the basic quality core

Execution

General

Key: recommendations_include_edit_quality

Starts on: 2023-05-22

Ends on: -

Team: DnA - PPP - Recommendation

Owner:

Experimentation Tool: Optimizely

Audience

Affected Locales:
ALL

Interaction Detection

Check

Traffic Interaction: Yes

Metric Interaction: Yes

Status: SUCCESS

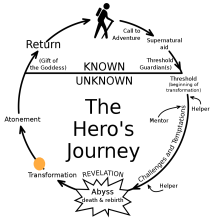
Last Update (UTC) on: 2023-05-23 12:51:01

Traffic Shared Experiments

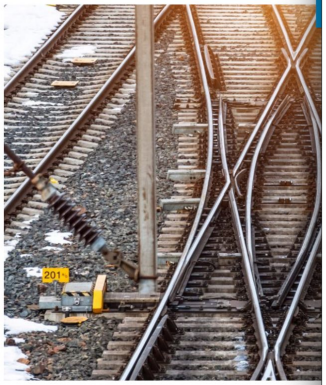
Experiment Y	Job Run Status	Traffic Interaction	Metric Interact...
na_-_studio_help_panel_experiment__ex...	SUCCESS	✗	✓
watch_retirement	SUCCESS	✗	✗

Actions

- Copy EXPO link
- Edit in Jira
- Configuration
- Results
- SRM



11. The Ultimate Boon

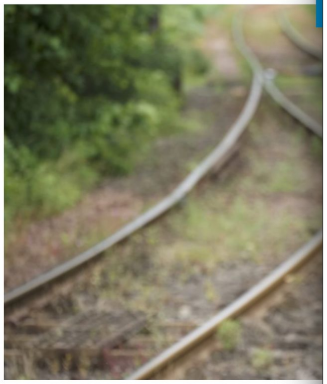


27 FEBRUARY 2023 BY LUKAS VERMEER
IN INNOVATION, EXPERIMENTATION

Interaction Effects Experimentation

Experiments allow us to test how ch behavior of our users. We make mar running at the same time. As we sca Vista, there will be an increased risk occurring.

We say an interaction has occurred experiments on a metric combined



4 APRIL 2023 BY LUKAS VERMEER
IN INNOVATION, EXPERIMENTATION

Avoiding Interactio Experimentation

In the previous post (find it here) in effects and discussed the potential undetected. In this post, we will dis interaction effects and share some conflict avoidance.

Some may think the ideal solution is simultaneously to avoid all possible

vista.io

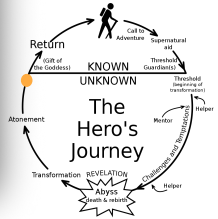
vista. DnA Data and Analytics

26 JUNE 2023 BY LUKAS VERMEER, ADRIANA APARICIO MARIJUAN
IN INNOVATION, EXPERIMENTATION

Detecting Interaction Effects in Online Experimentation

In the first two posts (one, two) in this series, we explained what interaction effects are and what their consequences could be if they remain undetected. We also listed several approaches to avoiding interaction effects and shared some of the tools and processes we use to enable conflict avoidance.

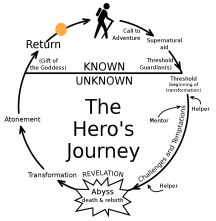
In this third and final installment in the series, we will discuss how to detect interactions and share code and tools we built at Vista to address this



15. The Crossing of the Return Threshold

Interaction effects are still not a problem worth worrying about, but people being worried about interaction effects is a problem worth worrying about.

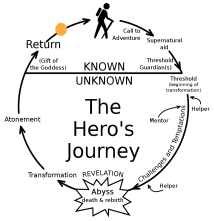
Avoid some, detect others. Regression is one approach to implement such detection.



17. Freedom to Live

Second Q&A

Your opportunity to ask me pretty much anything at all except for that one thing.



17. Freedom to Live