humu

Privacy, Triage, and Risk

Or, The Art of VP-Wrangling

Yonatan Zunger
Distinguished Engineer &
Chief Ethics Officer, Humu

zunger@humu.com

PEPR 2019

My Key Collaborators

- Lea Kissner – Privacy and respect in computing
- A. V. Flox – Psychological safety, activation, and trauma
- The Google Privacy Team – years of work and collaboration

The Talk I'm Not Giving: "Building for Trust"

- What is building for trust?
- Why build for trust?
- How do we build for trust?
- How do we assure ourselves we've build correctly?
- Where are the big challenges?

For the broad outline, see Lea Kissner's IWPE 2019 Keynote, or Sha Sundaram's PEPR talk from yesterday.

The Talk I'm Not Giving: "Building for Trust"

- What is building for trust?
- Why build for trust?
- **How do we build for trust?**
- How do we assure ourselves we've build correctly?
- Where are the big challenges?

For the broad outline, see Lea Kissner's IWPE 2019 Keynote, or Sha Sundaram's PEPR talk from yesterday.

How do we build for trust?

1. *Find all the parts.*
2. *Check the invariants.*
3. *Model the threats.*
4. *Triage and correct issues.*
5. *Document, document, document.*

How do we build for trust?

1. *Find all the parts.*
2. *Check the invariants.*
3. *Model the threats.*
4. ***Triage and correct issues.***
5. *Document, document, document.*

Triage includes three things

1. Identify possible responses to threats
2. Prioritize threat response (vs all other tasks)
3. Achieve consensus on prioritization

Triage includes three things

1. Identify possible responses to threats
2. Prioritize threat management (vs all other tasks)
3. **Achieve consensus on prioritization**

    Requires common language, alignment of understanding

    Good tools for (3) help (2), and vice-versa.

Why triage is hard

1. **Things people estimate poorly:**

   Rare, catastrophic events.

   Continuous invisible impacts that add up.

2. **People are drowning in priorities.**

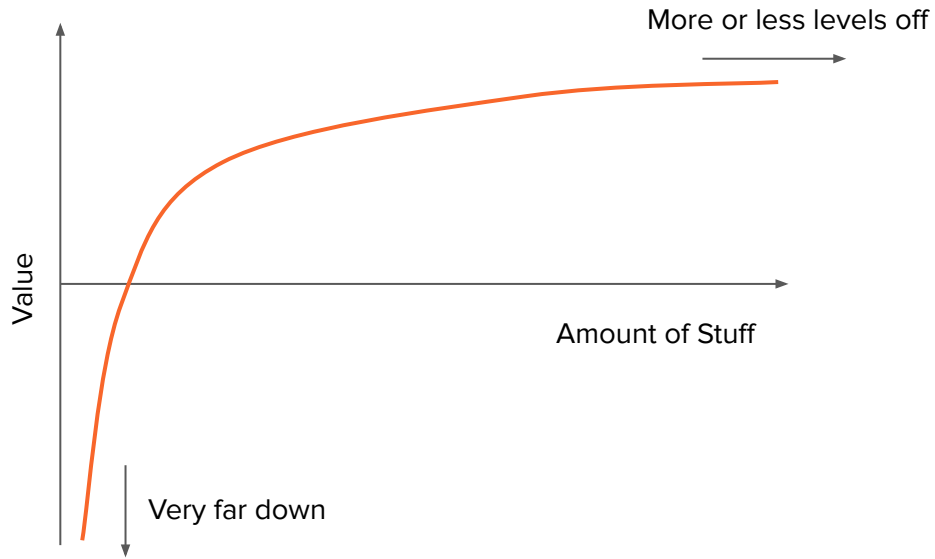
3. **Different people have different threat models.**

Four useful tools

1. *The Universal Value Curve*
2. *The Gottman Ratio*
3. *Sense vs. Presence*
4. *Activists vs. Operators*

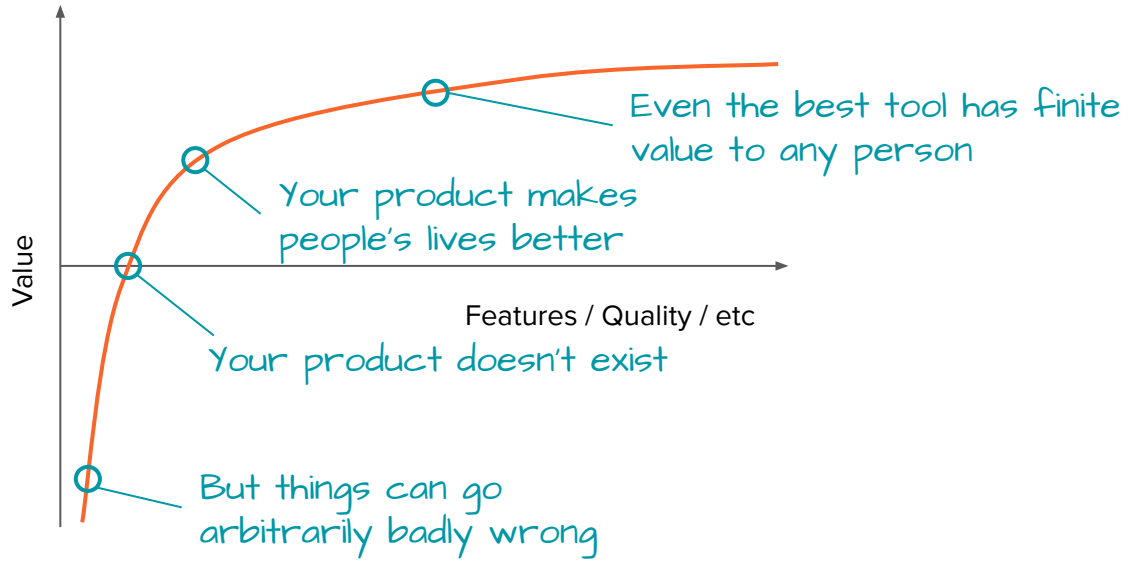
1 The Universal Value Curve

How do you trade product value vs. rare catastrophic risks?



1 The Universal Value Curve

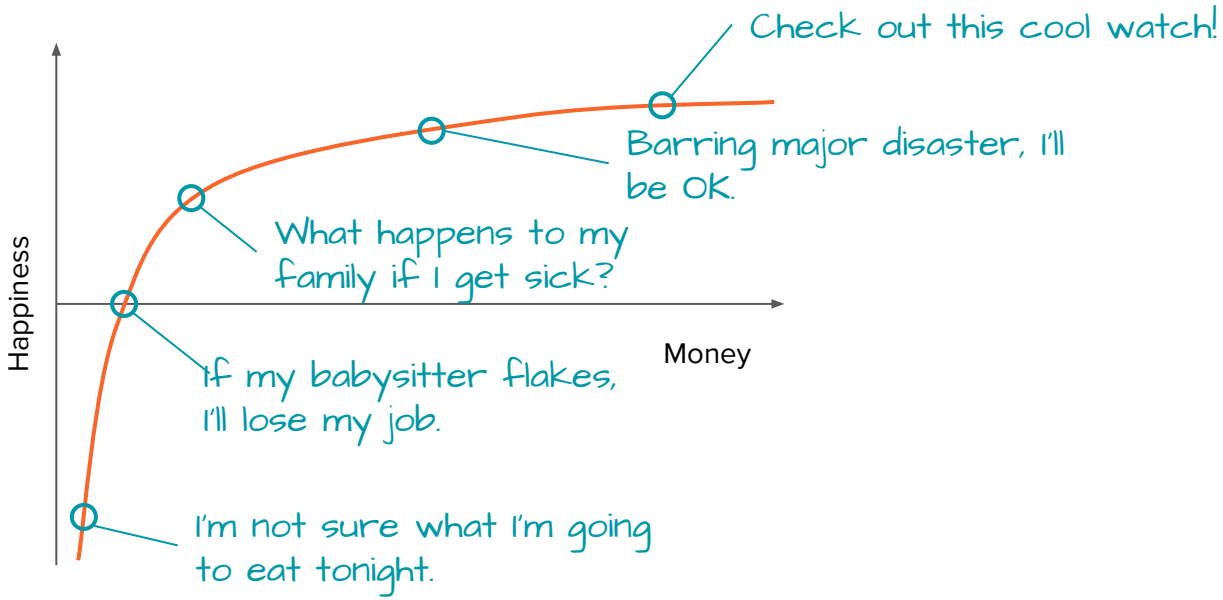
Applied to a single user of a specific product



1 The Universal Value Curve

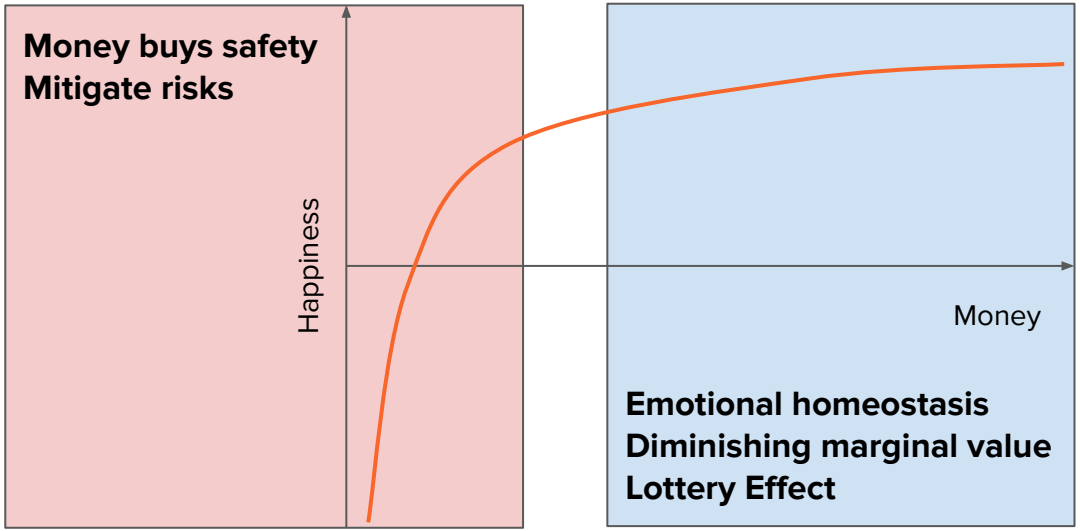
You've seen this curve before.

"Money can't buy you happiness, but poverty can buy you one hell of a lot of misery."



1 The Universal Value Curve

This also helps understand why this curve works

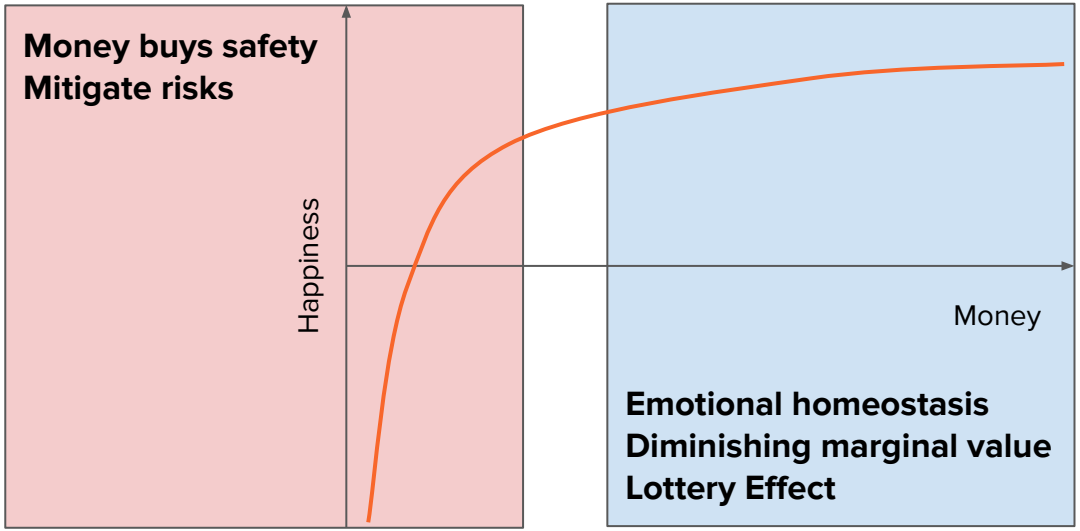


See also: Y. Zunger, "Financial Shock Wealth." NewCo Shift, Dec. 2017

See also: Notorious B.I.G., "Mo' Money, Mo' Problems," *Life After Death*

1 The Universal Value Curve

Most happiness comes from eliminating stressors

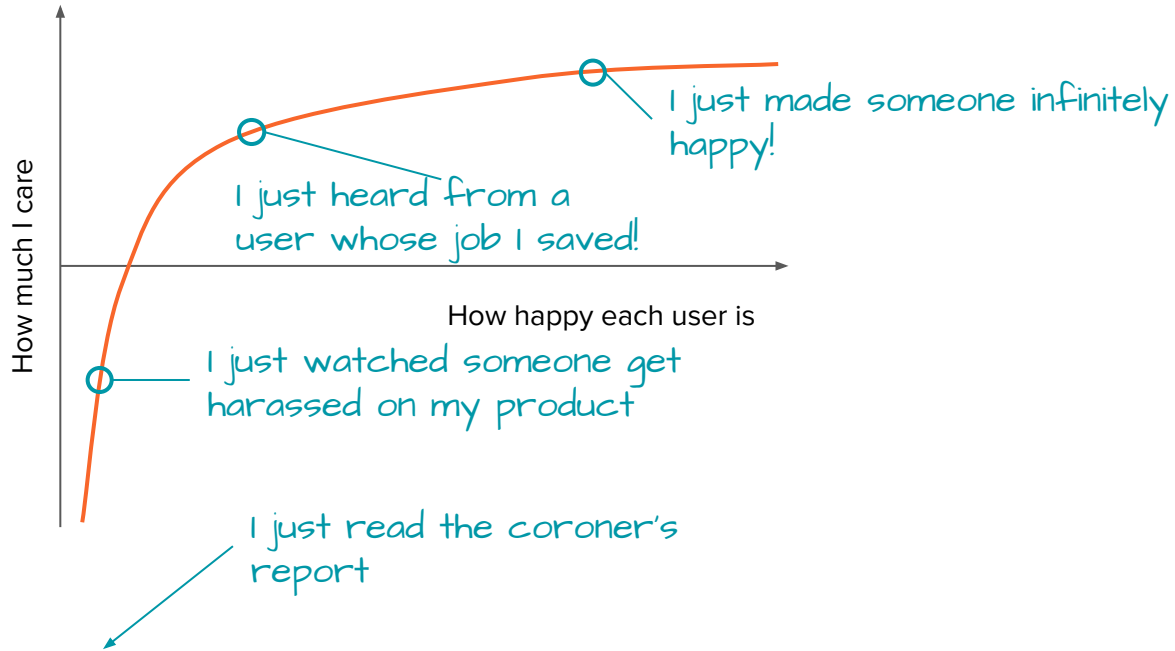


See also: Y. Zunger, "Financial Shock Wealth." NewCo Shift, Dec. 2017

See also: Notorious B.I.G., "Mo' Money, Mo' Problems," *Life After Death*

1 The Universal Value Curve

Applied to the total user population



1 The Universal Value Curve

That may seem obvious, but...

Do your product success metrics increase unboundedly per user?

Metrics like “total engagement-seconds” can create **perverse incentives**. Was the user any happier?

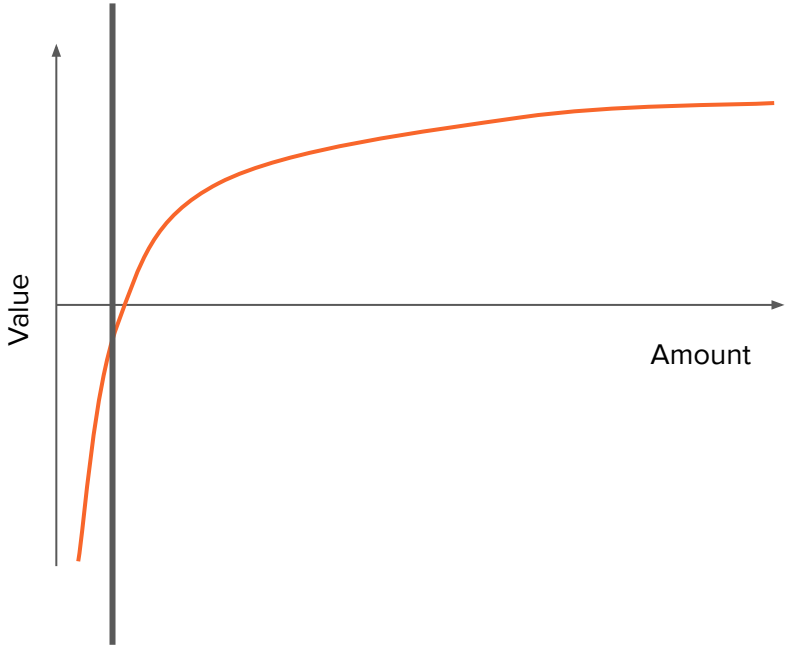
Do the numbers that drive your company align with user happiness? In the short term? In the long term?

1 The Universal Value Curve

Use this to **assess risk**, **design metrics**, and **convey tradeoffs**

Far LHS: Don't multiply.
bignum × smallnum = noise
You can't measure accurately enough.

Instead: Count, use narrative
Graph is *qualitative, illustrative*
"How many deaths per year is this feature worth?"



RHS: Be quantitative.
Examine unbounded metrics *carefully*.

Compare actively alienated users to users acquired.

Talk PR risk, legal risk, regulatory risk, human harm.

2 The Gottman Ratio

“Relationships need to have at least a 5 to 1 ratio of positivity to negativity during conflict”

Applied to user experience:

- Positive moments (“aha!”, task accomplished, pleasure) versus
- Negative moments (stuck, broken, harassment)

J. Gottman, *Why Marriages Succeed or Fail*, Simon & Schuster, 1995, ch. 1

2 The Gottman Ratio

What's actually happening?

- The body is designed to detect danger cues and go on alert (neuroception)
- "Negative moments" are **unsafety signals:** "did I just lose my file?" "Am I being attacked?"
  - Same kind of unsafety as in LHS of value curve!
  - Response via autonomous NS, 50-500msec; decay via meta-cognitive processes, seconds-hours.

J. Gottman, *The Science of Trust*, W. W. Norton, 2011, ch. 6

2 The Gottman Ratio

- As emotional resources deplete, the trigger threshold drops
  - People stop looking for safety and start confirming danger.
- Creates a **positive-feedback loop**:
  - Initial susceptibility $\propto$ person's baseline unsafety.
- A single source can trigger the signal many times
  - Re-reading alarming text
  - Encountering other people's danger responses
  - Attempt to remedy which fails

A. V. Flox, *Disrupting the Bystander*, Thorntree Press, 2019, ch. 3

2 The Gottman Ratio

Consequence: When people feel unsafe, they're…

- Less likely to explore, learn, exchange information
- Less likely to keep using the product
  - 30% of people who have experienced serious online harassment have stopped using an online service after witnessing harassment there.
    - 18% have experienced serious harassment
    - 41% of 18-to-29-year-olds

Y. Gong *et al.,* "Unfolding the Proactive Process for Creativity," *J. Mgt* 2012 38: 1611

M. Duggan, "Online Harassment 2017." Pew Research Center, Washington, D.C. (July 11, 2017)

2 The Gottman Ratio

Use this to **design interfaces**, **highlight user loss risk.**

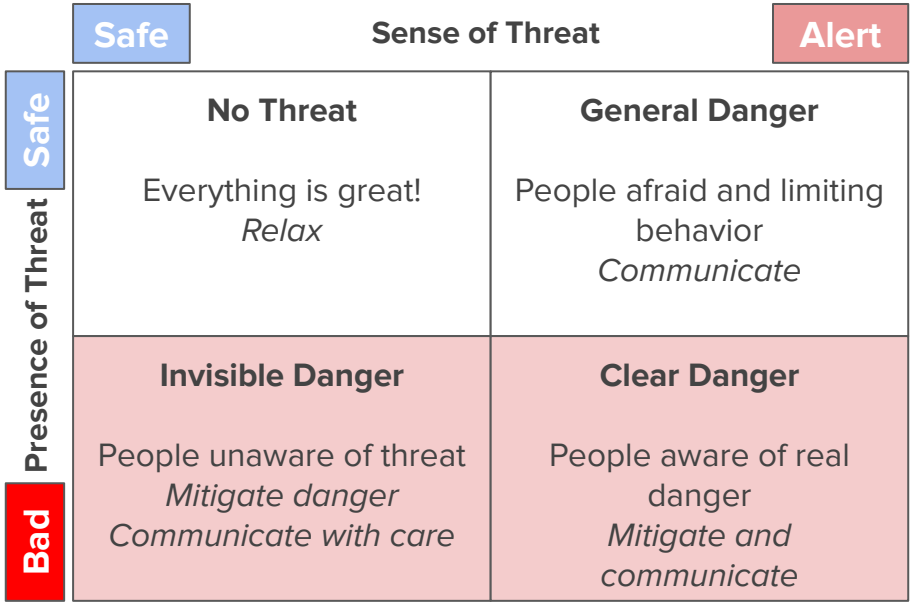
- Measure positive/negative ratio directly, when you can.
  - This can be a good user happiness metric
  - Gottman ratio hints at where it saturates
- When something goes wrong, minimize time until user can act to relieve the problem.
  - Easy report flow; quick help; get back to “safe state.”
  - Think 100s of msec, not minutes.

3 Sense vs. Presence

How do you prioritize user worries versus real dangers?

3 Sense vs. Presence

How do you prioritize user worries versus real dangers?

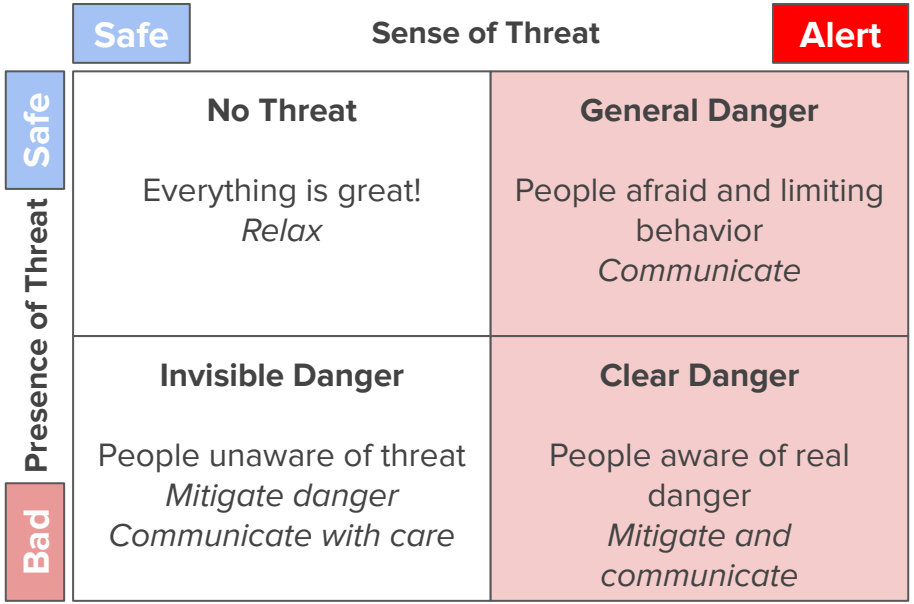


Traditional risk analysis

If this goes wrong, what happens? How do we prevent it? What do we communicate?

3 Sense vs. Presence

How do you prioritize user worries versus real dangers?



User perception affects action.

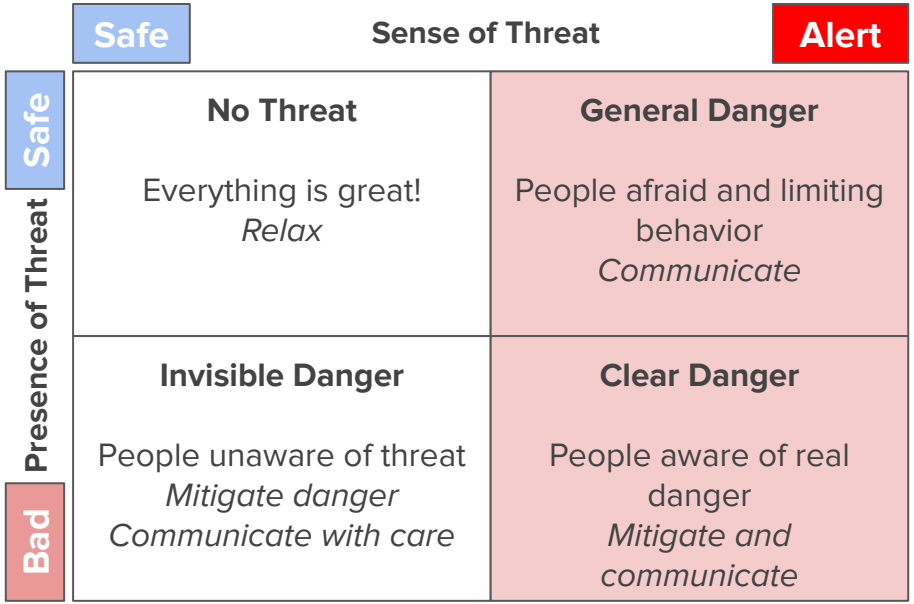
When people don't know, they assume the worst.

If you're legit, how do people know? How can they verify it?

Fear spreads in a vacuum

3 Sense vs. Presence

How do you prioritize user worries versus real dangers?



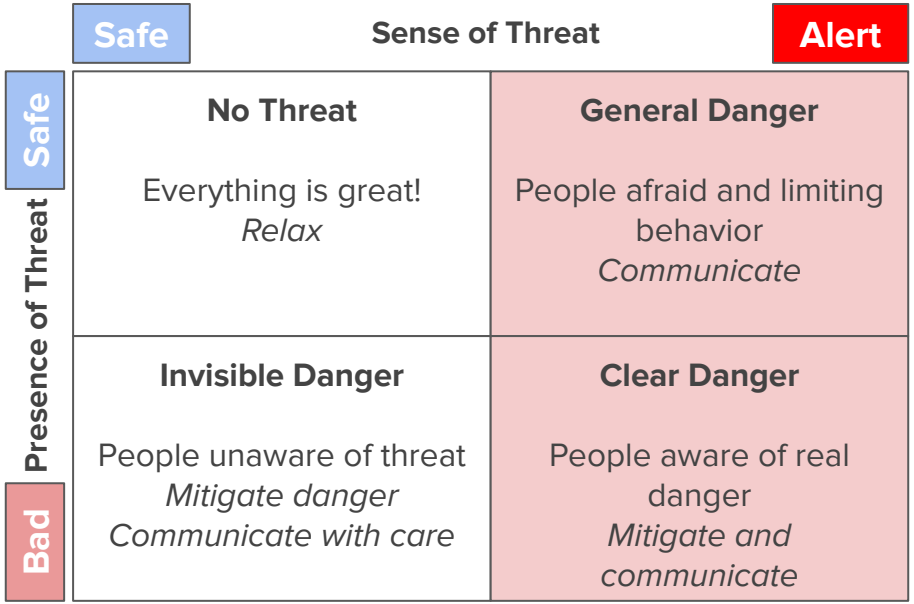
Beware of scared PR / legal / etc teams

One school believes “the less you say, the less risk.”

If people are already afraid, you’re already paying the price.

3 Sense vs. Presence

How do you prioritize user worries versus real dangers?



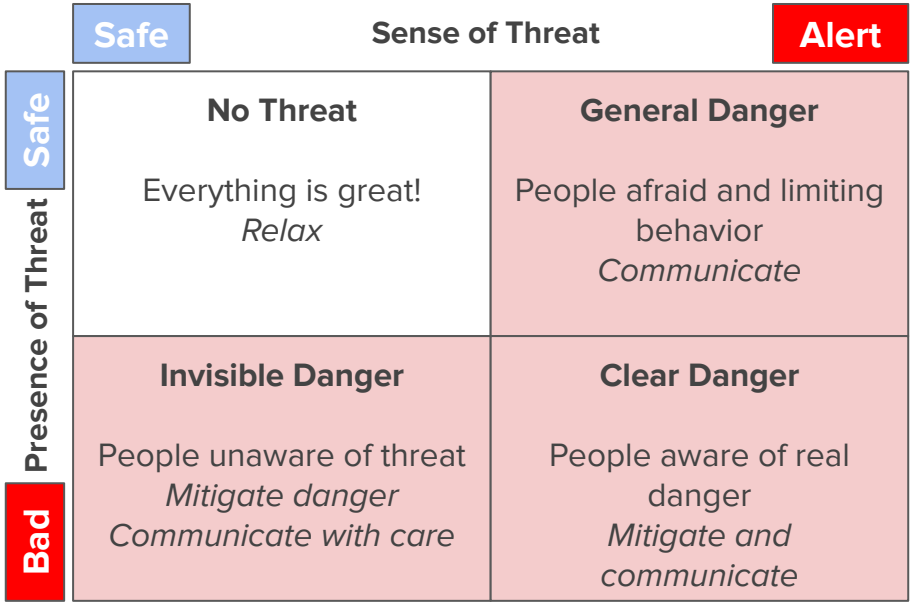
But look for unexpected allies!

PR, legal, marketing, or public policy teams often *want* to communicate.

They often fear the others will say no! Talk to everyone.

3 Sense vs. Presence

Use this to **talk about perception in the language of risk.**



If you're in the red, you have a problem.

The risk to your *company* is real in all three squares!

4 Activists vs. Operators

You’re worried about a risk; the product VP isn’t.

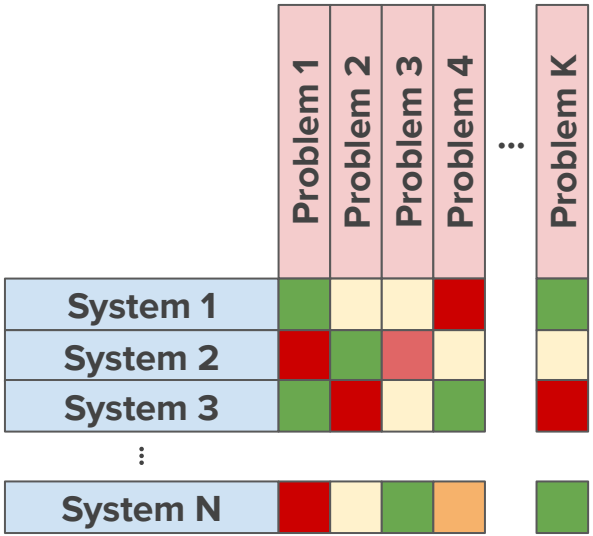
Are they right? Are they uninformed? Do they not care?

Activists view **systems** through the lens of a single **problem**.

System operators view **problems** through the lens of a single **system**.

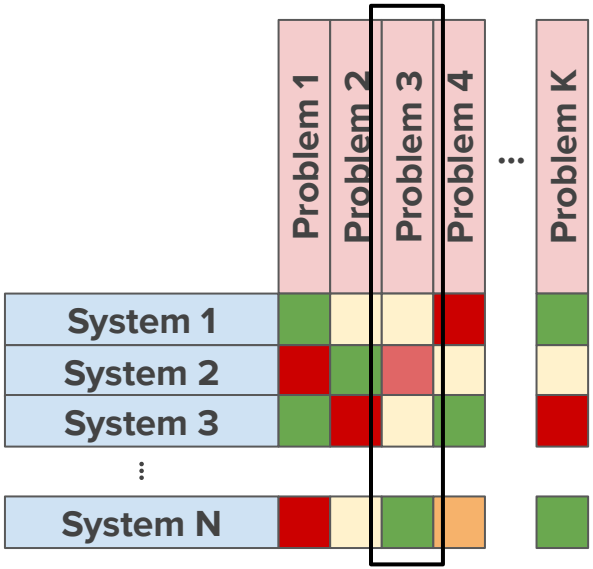
4 Activists vs. Operators

The system that most affects the problem ≠ The problem that most affects the system



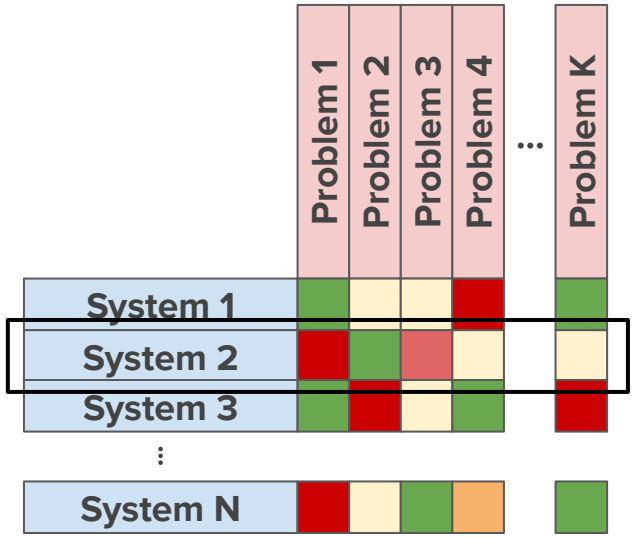
4 Activists vs. Operators

The system that most affects the problem ≠ The problem that most affects the system



4 Activists vs. Operators

The system that most affects the problem ≠ The problem that most affects the system



4 Activists vs. Operators

Sometimes priorities don’t align. When this happens, you often want something from an operator.

- You need to build social capital:
  - Do they have a persistent annoyance you can fix?
  - Can your proposal move the metrics they worry about?
  - If you can’t do it, can you get someone else to?

Look for *their* unspecified fears and pressures – and help fix them!

4 Activists vs. Operators

Caveat: The last bit applies to *non-adversarial* relationships when *you are not the person whose resources are being drained.*

- If basic goals don't align, this doesn't work well.
- If you're the injured party, you don't have the resources to do this.

But privacy professionals are often in this situation – Use it!

The Moral of the Story

Triage is both about estimating risks and convincing people.

The most useful tools give you a common language to do this.

Flox: “Remember the human”:

- People want to feel safe and be safe.
- That includes users and stakeholders alike.
- When you can help them do that, good things can flow.

A. V. Flox, *Disrupting the Bystander*, Thorntree Press, 2019