Hard Choices, Tight Timelines:

A Closer Look at Tradeoff Decisions during Incidents

SRECon Americas 2024

•••

Dr. Laura Maguire, Trace Cognitive Engineering/OSU
Courtney Nash, The VOID

What we talk about when we talk about tradeoffs.

Tradeoffs are…

- “Advantageous but conflicting properties (e.g., speed vs. accuracy) [that] are ubiquitous in cognition.” (Del Guice & Crispi, 2018)

Tradeoffs are…

- "Advantageous but conflicting properties (e.g., speed vs. accuracy) [that] are ubiquitous in cognition." (Del Guice & Crispi, 2018)

- "Choices between **different but interacting or conflicting goals**, between **values or costs** placed on different possible outcomes or courses of action, or between the **risks of different errors**" (Woods et al, 2006)

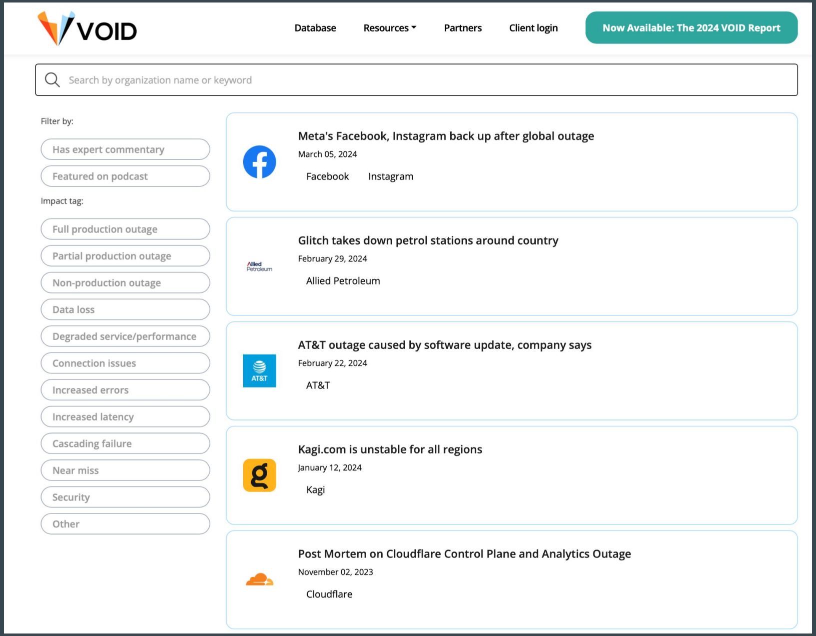
Tradeoffs are…

- “Advantageous but conflicting properties (e.g., speed vs. accuracy) [that] are ubiquitous in cognition.” (Del Guice & Crispi, 2018)

- “Choices between **different but interacting or conflicting goals**, between **values or costs** placed on different possible outcomes or courses of action, or between the **risks of different errors**” (Woods et al, 2006)
    - “while facing uncertainty, risk, and the pressure of limited resources (e.g., time pressure; opportunity costs).”

How we investigated tradeoffs in incident response.

Incident Data in the Wild: The VOID



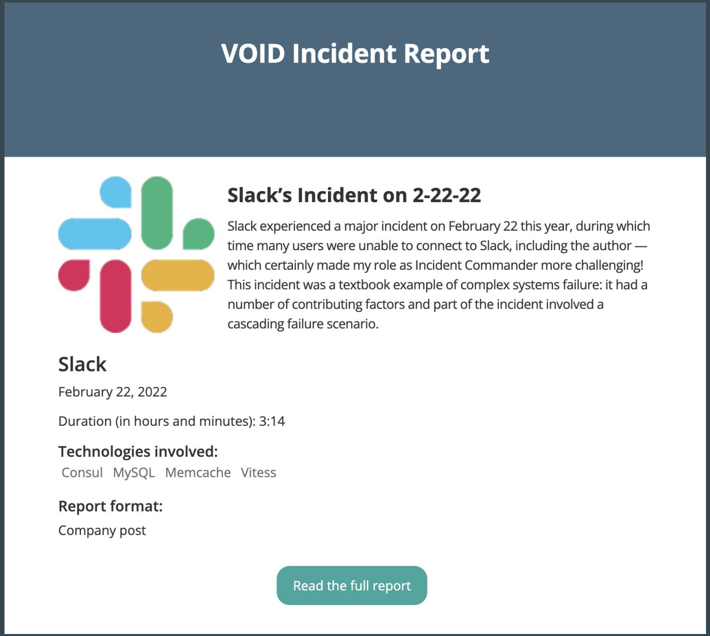
Meta's Facebook, Instagram back up after global outage
March 05, 2024
Facebook Instagram

Glitch takes down petrol stations around country
February 29, 2024
Allied Petroleum

AT&T outage caused by software update, company says
February 22, 2024
AT&T

Kagi.com is unstable for all regions
January 12, 2024
Kagi

Post Mortem on Cloudflare Control Plane and Analytics Outage
November 02, 2023
Cloudflare



VOID Incident Report

Slack's Incident on 2-22-22

Slack experienced a major incident on February 22 this year, during which time many users were unable to connect to Slack, including the author — which certainly made my role as Incident Commander more challenging! This incident was a textbook example of complex systems failure: it had a number of contributing factors and part of the incident involved a cascading failure scenario.

Slack

February 22, 2022

Duration (in hours and minutes): 3:14

Technologies involved:
Consul MySQL Memcache Vitess

Report format:
Company post

https://thevoid.community

What’s in The VOID?

10k+ public incident reports from nearly 600 organizations, from 2008 up to present day.

In a variety of formats:

- Social media posts
- Status pages
- Blog posts
- Conference talks
- News articles
- Tweets
- Comprehensive retrospectives/postmortem reports

Metadata including:

- Organization
- Date of incident
- Date of report
- Report type
- Duration
- Technologies involved
- Impact type
- Analysis format
- Severity

Narrowing the search space

- Tradeoff: 2
- Sacrifice: 2
- Rolling back/reverting:
  - Rollback: 60
  - Roll back: 10
  - Revert: 70
  - Reverted: 45
- Disabling a feature
  - Disable: 101
  - Disabled: 65
- Potential data loss
  - Data loss: 18
  - Restoring backup: 2

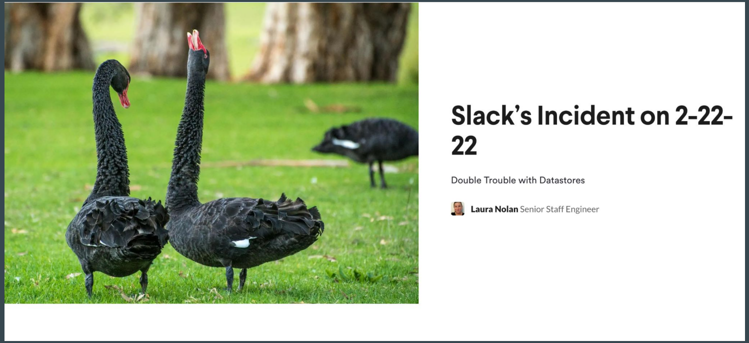
A Few Examples

A few likely trade off decision examples did pop up:

- Slack's Incident on 2/22/22
- Facebook 2021 outage
- Datadog Multi-Region Infrastructure Connectivity Issue
- Reddit Pi Day outage

Slack's 2/22/22 Incident

"These slow requests were causing resource exhaustion in our database tier and were preventing other requests—from users who had booted clients—from succeeding. Therefore we made a decision to throttle client boot requests. We knew that t**his** throttling would mean that **users without booted clients would be unlikely to be able to connect to Slack** — but **the tradeoff was that users who did have booted clients would likely see relatively normal service restored**. Furthermore, reducing load would reduce the number of database queries timing out, and thus allow the cache to fill." —Laura Nolan, Senior Staff Engineer



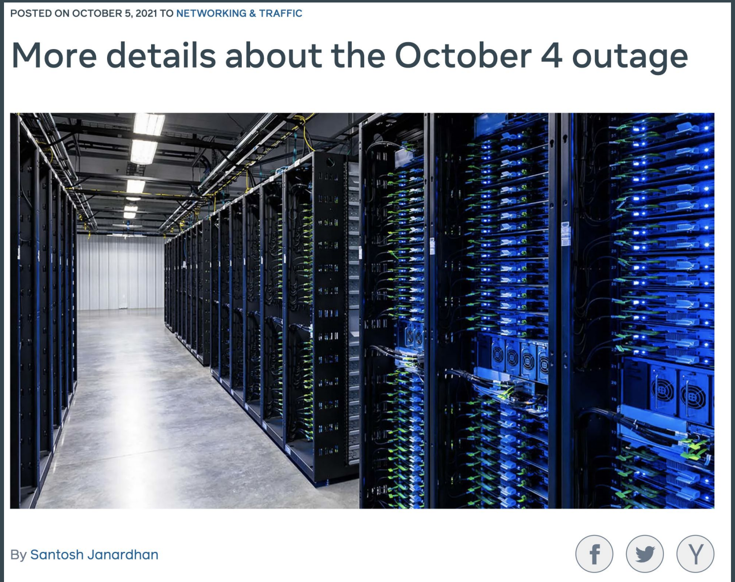
Slack's Incident on 2-22-22

Double Trouble with Datastores

Laura Nolan Senior Staff Engineer

Meta/Facebook's 10/4/2021 Incident

“We've done extensive work hardening our systems to prevent unauthorized access, and it was interesting to see how that **hardening slowed us down** as we tried to recover from an outage.... I believe **a tradeoff like this is worth it — greatly increased day-to-day security vs. a slower recovery from a hopefully rare event like this.**” —Santosh Janardhan, Head of Infrastructure



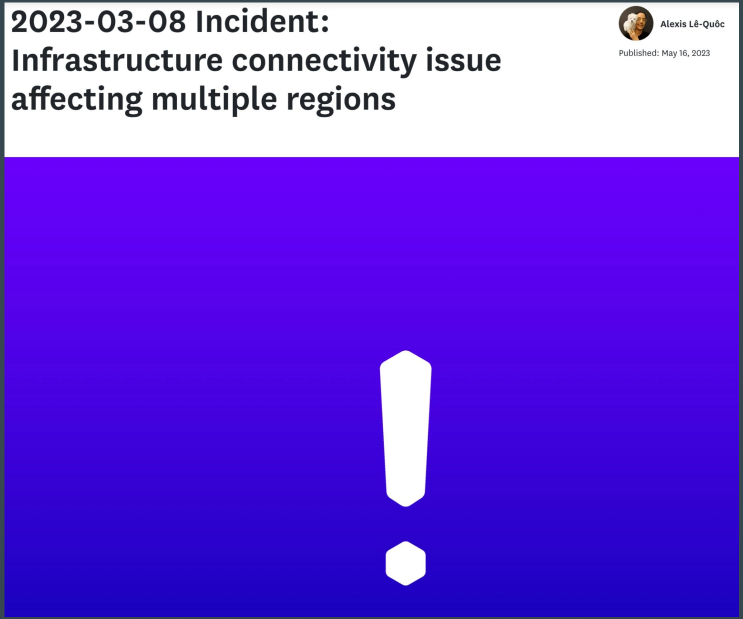
POSTED ON OCTOBER 5, 2021 TO NETWORKING & TRAFFIC

More details about the October 4 outage

By Santosh Janardhan

Datadog’s 03/08/23 Incident

“In all cases, our **number one priority** was to restore the processing of live data...Most important, **usable live data and alerts are much more valuable than access to historical data**. And even among all the live data, **data that is actively monitored or visible on dashboards is more valuab**le than the rest of live data. We will take this clear hierarchy into account in how we handle processing and access in degraded mode...**this may take the form of having only urgent data accessible and processed in degraded mode.**” —Alexis Le-Quoc, CTO



2023-03-08 Incident: Infrastructure connectivity issue affecting multiple regions

Alexis Lê-Quôc

Published: May 16, 2023

Reddit’s Pi Day Incident

“We were running low on constructive ideas, and the outage had gone on for over two hours at this point. **It was time to make the hard call; we would make the restore from backup**. Knowing that most of the worker nodes we had running would be invalidated by the restore anyway, we started terminating all of them, **so we wouldn’t have to deal with the long reconciliation after the control plane was back up**.”
—grumpimusprime, Compute team

r/RedditEng · 1 yr. ago
grumpimusprime ADMIN

You Broke Reddit: The Pi-Day Outage

Cute error image friends, we love them.

VOID Results

Organizations don’t tend to discuss/present tradeoff decisions in public incident reports.

There are tradeoffs about capturing tradeoffs in public reports

1. Purposes of internal vs external incident reports
2. These discussions ARE happening internally
3. Sharing these discussions publicly can help normalize the fact that these types of tradeoff decisions are inevitable within complex systems

The problem with asking “how do you make tradeoff decisions?”

Vignette methods

“Vignettes are short descriptions of a scenario for which participants are required to make a decision.

Through analysing the information within a scenario from the perspective of one’s **knowledge and experience**, they aim to **simulate the mental processes** of participants for making real and complex decisions.” (Reader et al, 2018)

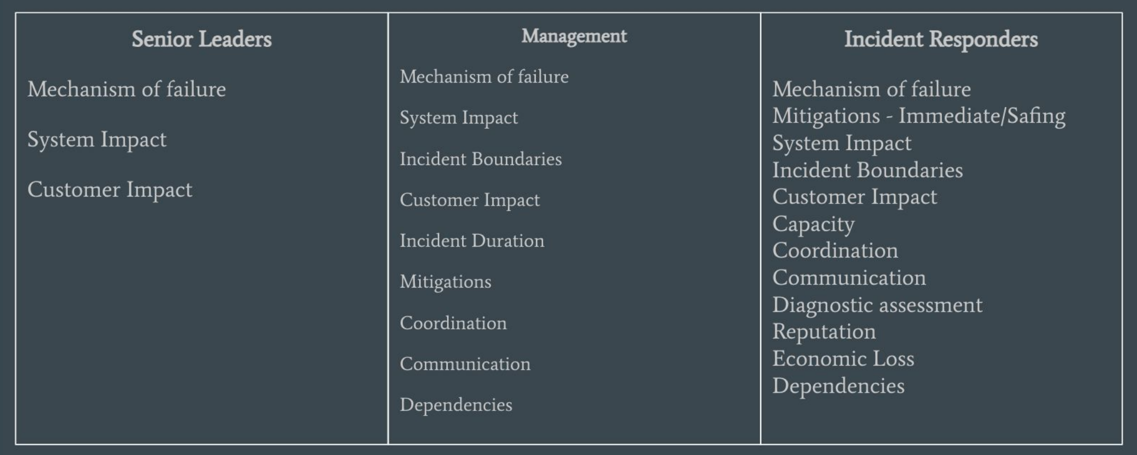
“Vignettes have been used to elicit **cultural norms** derived from respondents’ **attitudes and beliefs** about a specific situation.” (Barter & Renold, 1999)

Data collection & analysis - how do we ‘pick people’s brains’?

-Cognitive probes designed to elicit thought processes about trade off decisions

-Fielded to Individual Contributors, Managers, and Senior Leaders

-Detailed thematic analysis



Senior Leaders	Management	Incident Responders
Mechanism of failure System Impact Customer Impact	Mechanism of failure System Impact Incident Boundaries Customer Impact Incident Duration Mitigations Coordination Communication Dependencies	Mechanism of failure Mitigations - Immediate/Safing System Impact Incident Boundaries Customer Impact Capacity Coordination Communication Diagnostic assessment Reputation Economic Loss Dependencies

Thematic Analysis

- Establish research question related to the topic and ask those questions
- Assign *codes* to each relevant item of text from answers
- Collect codes into *themes*
- Each theme captures a prominent aspect of the data in a patterned way
- Revisit the themes in relationship to the research question

The incident vignette

The incident progresses

A further cascading effect...

The probes

- Considerations across time
- Recruitment
- Engagement of relevant parties
- Mitigating and minimizing risks
- The effects of increasing uncertainty or unexpected events

Participant demographics

N = 27

Distribution of respondents

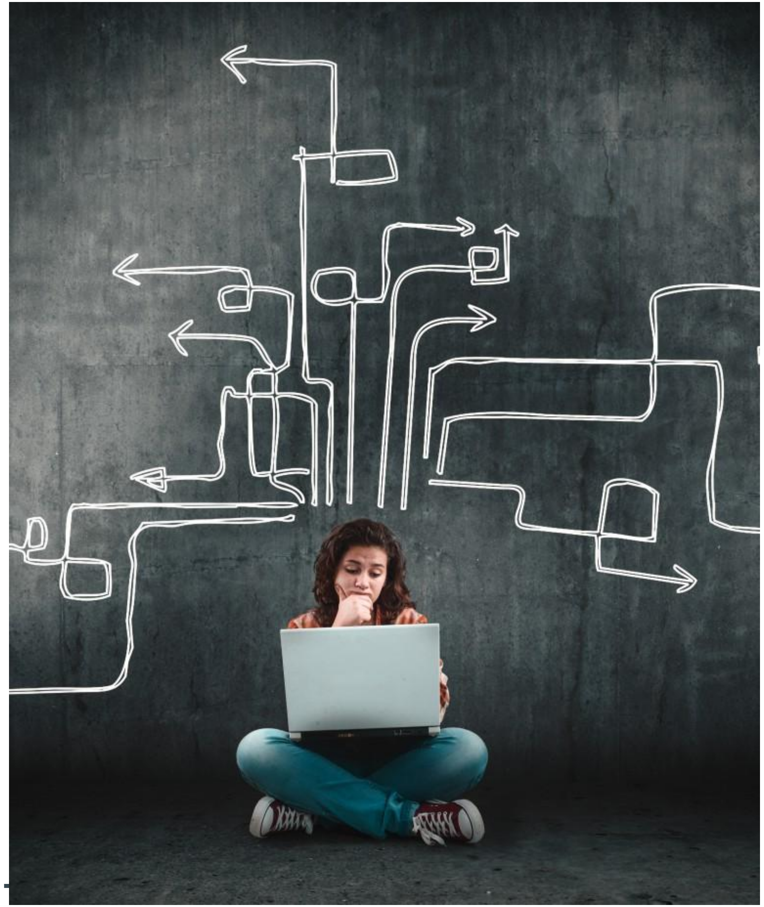
- 16% Senior Leadership
- 20% Manager or Skip level manager
- 64% Individual Contributors

Distribution

- Answered independently
- As a pairing

What we found

1. Tradeoff decision making in incidents is complex.



1. Tradeoff decision making in incidents is complex.

Senior Leaders	Management	Incident Responders
Mechanism of failure System Impact Customer Impact Incident Support - Coordination Clarity of Problem Definition Impacted Party Engagement Communication Coordination Compliance Capability Customer Impact - Actual/Potential Response Time Legal Implications	Mechanism of failure System Impact Incident Boundaries Customer Impact Incident Duration Mitigations Coordination Communication Dependencies	Mechanism of failure Mitigations - Immediate/Safing System Impact Incident Boundaries Customer Impact Capacity Coordination Communication Diagnostic assessment Reputation Economic Loss Dependencies

1. Tradeoff decision making in incidents is complex.

Tradeoffs decisions are technical, organizational, *and* social.

1. Tradeoff decision making in incidents is complex.

“It made me a little queasy thinking about being involved in something like this. In a very long career, it's only happened maybe twice where I was personally concerned about liability. It's not fun, and nobody prepares you for it when they add you to the pager rotation.”

1. Tradeoff decision making in incidents is complex.

“It made me a little queasy thinking about being involved in something like this. In a very long career, it's only happened maybe twice where I was personally concerned about liability. It's not fun, and nobody prepares you for it when they add you to the pager rotation.”

Responder reacting to the revelation that downstream services still had access to the unredacted information and were using it in breach of the law:

1. Tradeoff decision making in incidents is complex.

“It made me a little queasy thinking about being involved in something like this. In a very long career, it's only happened maybe twice where I was personally concerned about liability. It's not fun, and nobody prepares you for it when they add you to the pager rotation.”

Responder reacting to the revelation that downstream services still had access to the unredacted information and were using it in breach of the law:

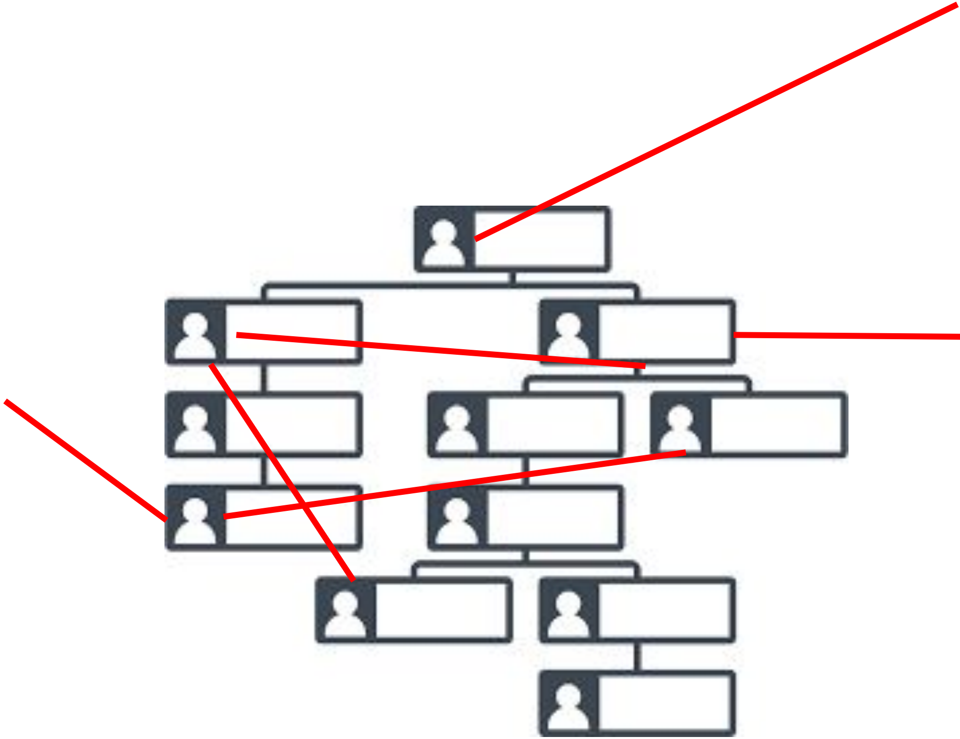
*"Oh f*ck"*

2. Tradeoff decisions are considered and managed differently across roles and levels within the organizations

2. Tradeoff decisions are considered and managed differently across roles and levels within the organizations

Senior Leaders	Management	Incident Responders
Mechanism of failure System Impact Customer Impact	Mechanism of failure System Impact Incident Boundaries Customer Impact Incident Duration Mitigations Coordination Communication Dependencies	Mechanism of failure Mitigations - Immediate/Safing System Impact Incident Boundaries Customer Impact Capacity Coordination Communication Diagnostic assessment Reputation Economic Loss Dependencies

3. Tradeoff decisions cross boundaries



3. Tradeoff decisions cross boundaries

“More people involved muddies decision making sometimes, slowing things down, but having the necessary teams involved means that decisions made will have a higher chance of considering all relevant and important information.”

3. Tradeoff decisions cross boundaries

"More people involved muddies decision making sometimes, slowing things down, but having the necessary teams involved means that decisions made will have a higher chance of considering all relevant and important information."

"Legal is a tough one to involve, as they may grind things to a halt, but it may also help move things forward if managed by an incident commander as they can scope the questions for a lawyer in a way that will allow the team to make decisions around the regulatory changes."

3. Tradeoff decisions cross boundaries

"involving people could entirely change the direction and requirements of the response."

3. Tradeoff decisions cross boundaries

“involving people could entirely change the direction and requirements of the response.”

“executive leadership: kept in the loop, probably not in the room, unless they are there to support legal”

4. Knowing more about organizational context increases focus on anticipation and optimization for others.

4. Knowing more about organizational context increases focus on anticipation and optimization for others.

*"With respect to legal, **I would not expect them to understand the technical details** of what is happening. Most likely **I would not pull them into an incident channel** or incident bridge, unless I knew that the person I was dealing with was particularly technical and good in such situations (i.e. knew how to behave and not be disruptive by asking a lot of questions or trying to take charge). If I were incident commander, **I would very likely communicate with the legal advisor privately, rather than in an open channel.**"*

4. Knowing more about organizational context increases focus on anticipation and optimization for others.

“Senior management can create a chilling effect, and can confuse authority in an incident. I'd quite likely opt to keep management out of the technical incident response and discuss the issue and the options privately.”

4. Knowing more about organizational context increases focus on anticipation and optimization for others.

“That there was a bug in a change that seemed to have been left until the last minute is very typical in our industry ... released the same day as the law changed, with a mere two weeks for sanity checks? that's a single sprint most places, not much will get done if it wasn't planned work.”

“Throughout this whole process I'd be feeling more and more frustrated with whatever dysfunction led to this data dependency problem, and I would have more and more tendency towards statements like "this wouldn't be a problem if [other people] could just do their jobs".

5. Costs and benefits of tradeoffs may be unevenly distributed.

5. Costs and benefits may be unevenly distributed.

Senior Leaders	Management	Incident Responders
IST AAT, Le IC —-------------- Co PO-IS, PO-AS CS PR Sa Fi EM	IST AS —------------- EM Le CS SL IC	IST AAT, Le Co —--------------- PO-IS, PO-AS SL CS PR Sa IC Fi EM

IST-Impacted System Team; AAT- All Affected Teams, PO- Product Owner, AS/IS-Affected System/Impacted System, Le-Legal; Co-Compliance; PR-Public Relations/Communications, IC - Incident Command, EM - Engineering Management, SL - Senior Leader, Sa- Sales, Ma-Marketing, Fi - Finance

6. Tradeoff decisions evolve over time.

$$\frac{\Delta}{t}$$

6. Tradeoff decisions evolve over time.

“This feels like we need to first limit the impact by creating a band-aid solution and then modify the core service to migrate to a newer endpoint”

6. Tradeoff decisions evolve over time.

As the incident progressed...

- and more information becomes available, increased willingness to bring more roles in.
- the range of ‘stop gap’ mitigations expanded
  - *“pay fines for a while and report the violations to regulators”*
  - *“immediately divert resources to rapidly removing as much dependency on the data as possible, bringing resources together to do it rapidly in the incident context. For the services that can't be rapidly upgraded, I'd explore options to deliver synthetic data in place of the illegal payloads.”*
  - *“Get business approval to expedite consulting resources as needed”*

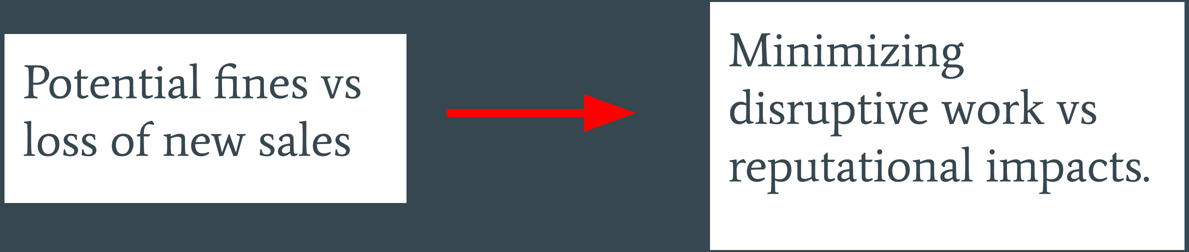
6. Tradeoff decisions evolve over time.

- The issue going public showed very little differentiation between the levels.
- Most said their consideration of the issue was treated seriously because of the regulatory violation.
    - *"I think the legal violation is where my considerations changed, vs. when it became public."*
    - *"It shifts some of my focus from containment to transparency. We still want to make sure we retain information about what happened when it was not public, and maybe even more now we need to prevent anyone from trying to hide that as there may be obstruction liability."*

7. Some goals & priorities get trashed along the way.

7. Some goals & priorities get trashed along the way.

- As awareness of the extent of the problem grows, the emphasis on economic loss shifts to company impact and reputation.



Potential fines vs loss of new sales

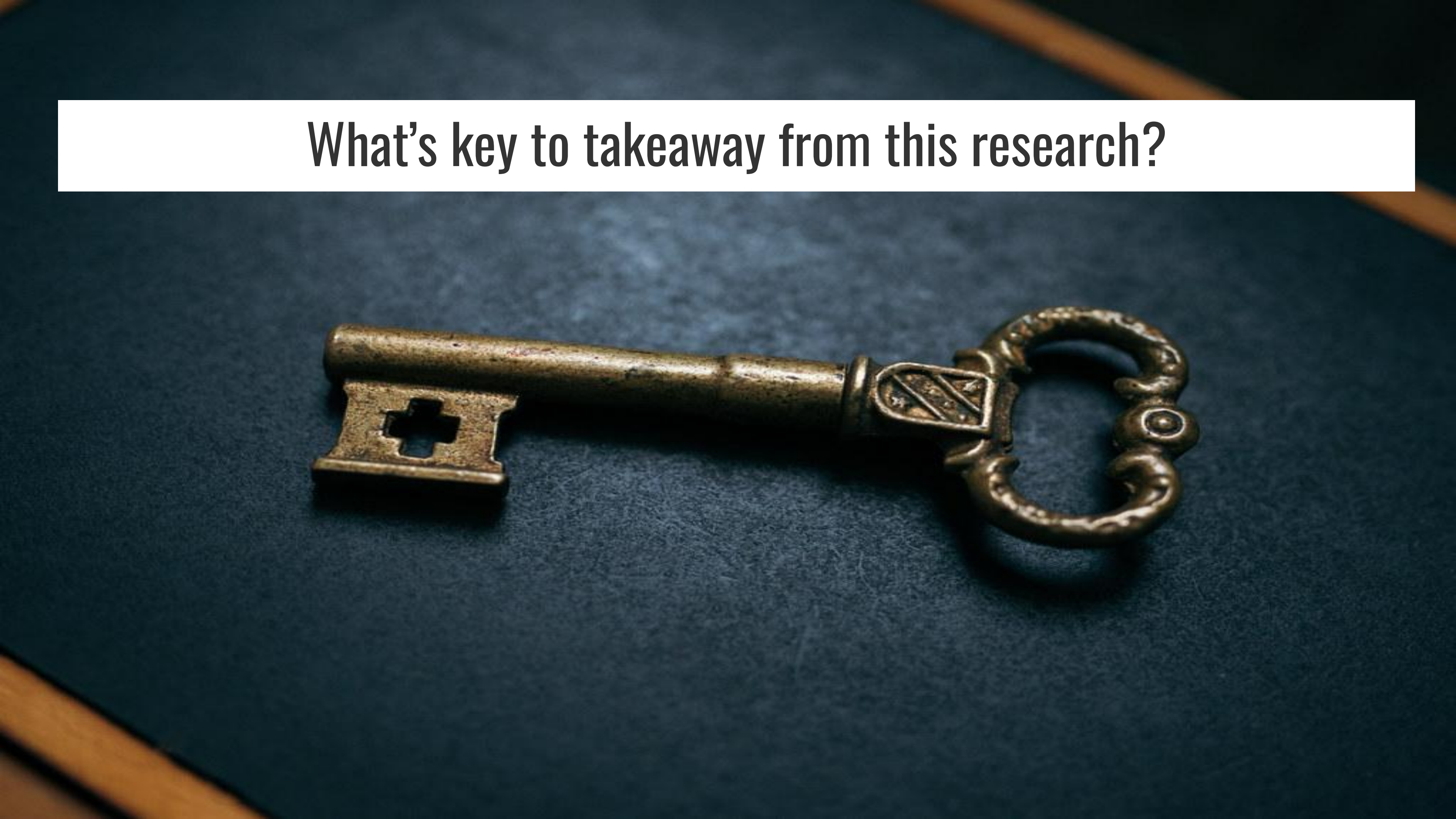
Minimizing disruptive work vs reputational impacts.

7. Some goals & priorities get trashed along the way.

“Cost/tradeoff is how much work it takes to query accurately (shortcuts, alternative querying patterns) vs. remediation (reducing data corruption and having existing expected query patterns to match their expected results)”

“Are we going to be on the front page of hacker news? CNN?”

What’s key to takeaway from this research?



Takeaways

1. Making tradeoff decisions can be as complex as the technical debugging. Let's start recognizing this and developing these skills.
2. Tradeoff decisions are managed differently across the organization. Bringing those perspectives to bear effectively takes practice.
3. Invest in cross boundary decision-making capabilities
4. Encourage decision-making that emphasizes anticipation and optimization across boundaries
5. Be transparent about costs and benefits and ask if they align with the values and long-term success of the organization.

Takeaways

6. Tradeoff decisions evolve over time, so practice effectively reframing the problem and continual model updating to avoid frustrations and oversimplifications.

7. Recognize when and how conditions are changing in ways that requires some goals and priorities to be trashed. Be explicit so others can adapt to this reality.

Limitations

- Self selection
- Self identification of level
- Variability in role titles and authority across organizations
- Duration of the vignette may have impeded more senior leadership participation

WHAT'S
NEXT?

What’s next? Future research

- Collect more data!
- Map the **extent of the information needed** for trade off decisions
- Better understand **role goals and priorities** in organizations
- Evaluate the effects of introducing **trade off decision debriefing in incident reviews**.

Acknowledgements

Fred Hebert
Staff Site Reliability Engineer
Honeycomb.io

Stay in touch!

Laura Maguire info@tracecognitive.com

Courtney Nash courtney@prowler.com

References

Del Giudice M, Crespi BJ. Basic functional trade-offs in cognition: An integrative framework. Cognition. 2018 Oct;179:56-70. doi: 10.1016/j.cognition.2018.06.008. Epub 2018 Jun 15. PMID: 29909281.

Barter, C., & Renold, E. (1999). The use of vignettes in qualitative research. Social research update, 25(9), 1-6.

Reader, T. W., Reddy, G., & Brett, S. J. (2018). Impossible decision? An investigation of risk trade-offs in the intensive care unit. Ergonomics, 61(1), 122-133.

Woods, D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2017). Behind human error. CRC Press.

www.tracecognitive.com/vignette

https://thevoid.community

https://u.osu.edu/csel/