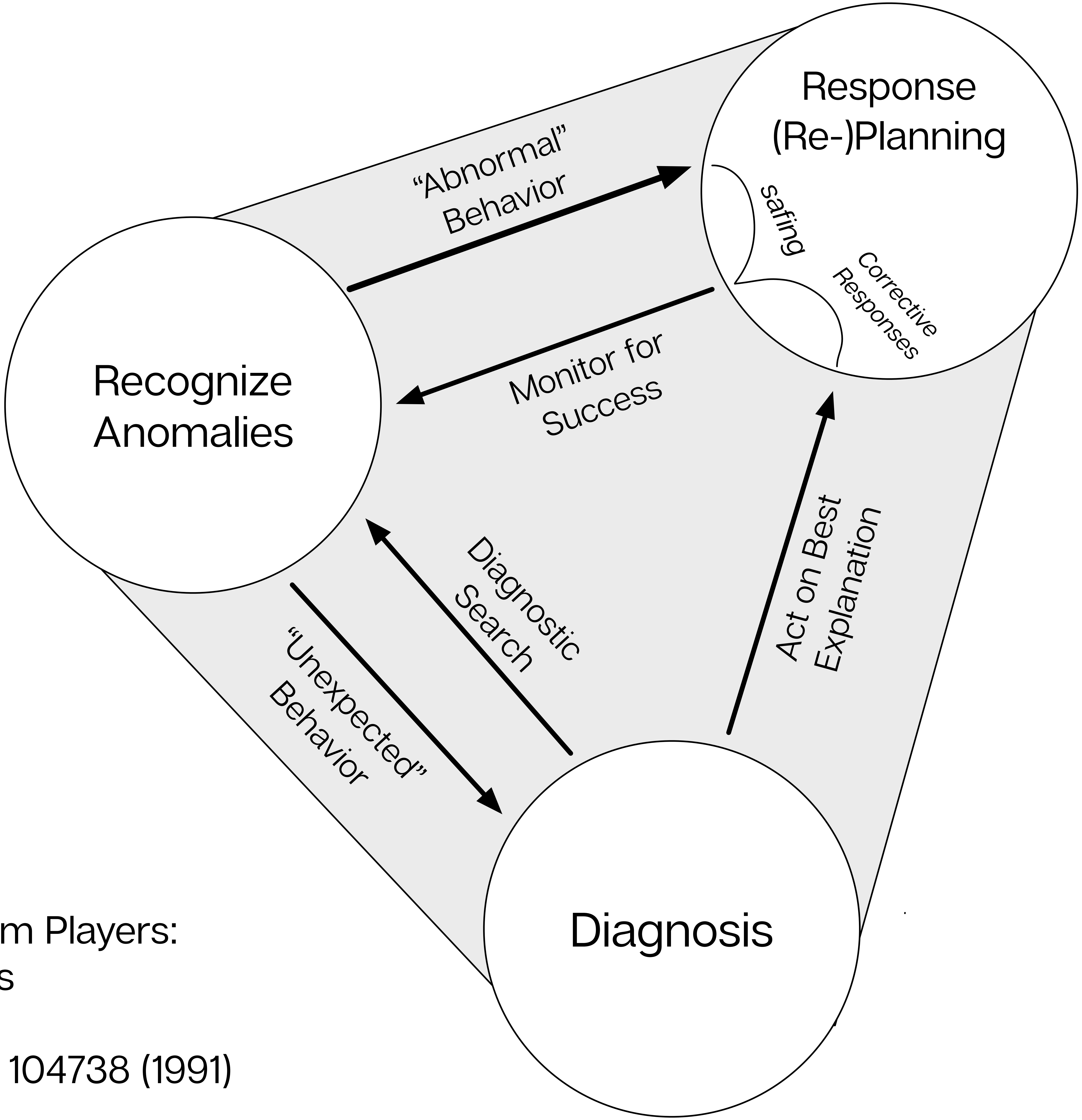
Hard Problems We Handle in Incidents

...but aren't often recognized

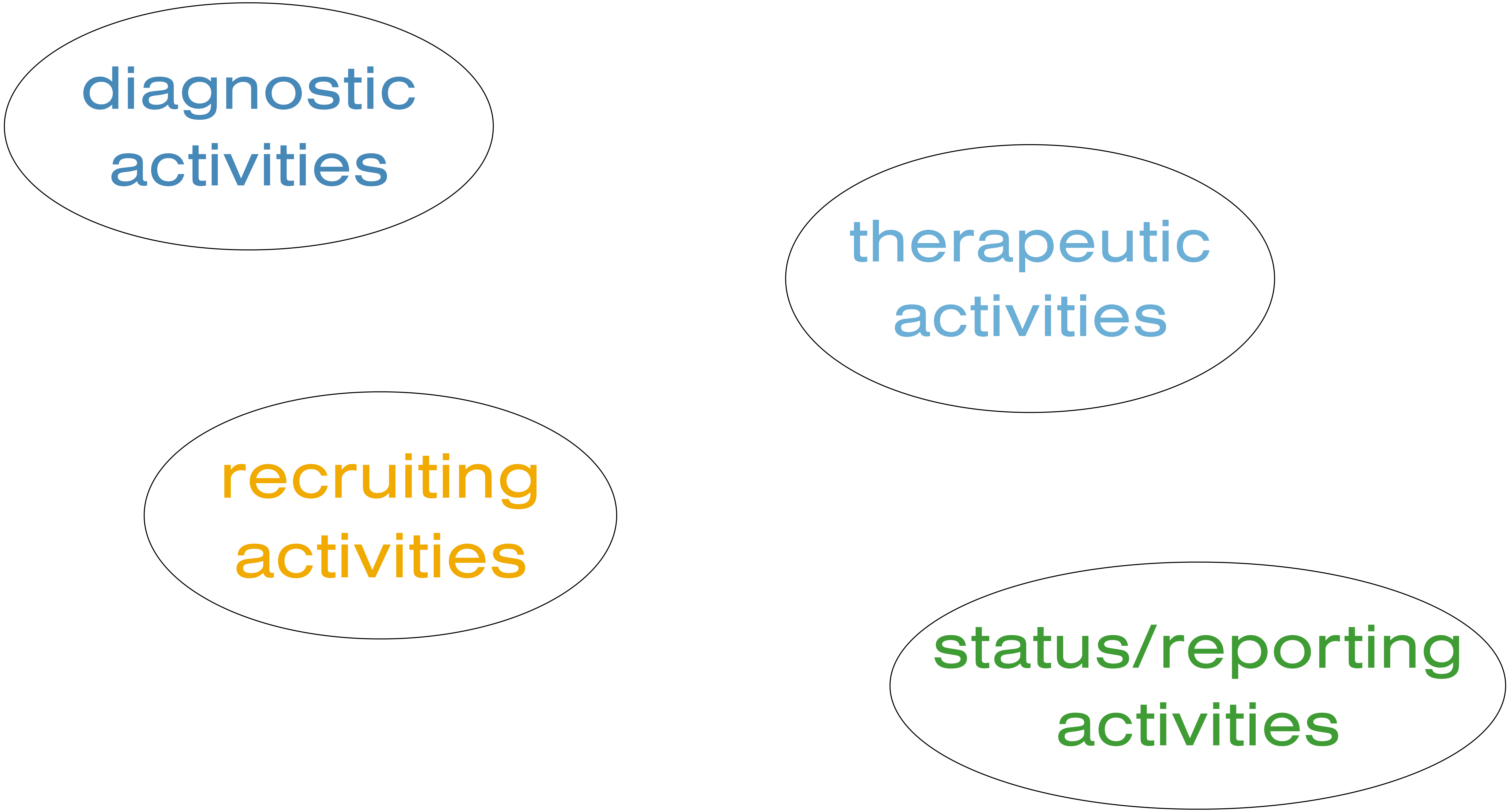
John Allspaw
Adaptive Capacity Labs

incidents are
bigger on the inside



Making Intelligent Systems Team Players:
Case Studies and Design Issues

NASA Technical Memorandum 104738 (1991)



diagnostic activities

therapeutic activities

recruiting activities

status/reporting activities

diagnostic activities

Observations & Signals

What is happening?
How is it happening?
How did it get like this?
What will it do next?

What **tools**...

- could I use?
- are others already using?

What **observations**...

- should I share with others?
- do I need to explain? how much detail?

How much **attention** should I pay...

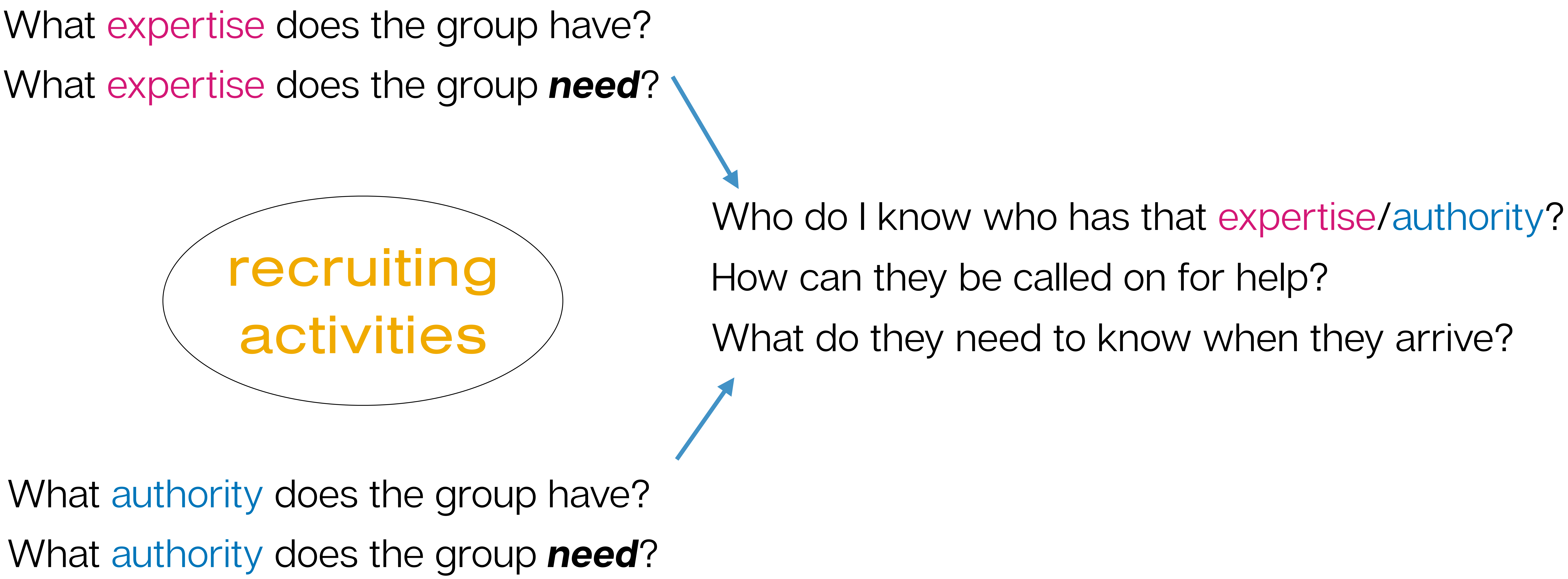
- to what *I'm* doing and seeing?
- to what *others* are doing and seeing?

What *can* we do?
What *are we able* to do?

- ...to lessen the impact, or prevent it from getting worse?
- ...to halt/revert systems, sacrificing potential data?
- ...to resolve the issue entirely?

therapeutic activities

- What options can ***I*** see?
- What options are ***others*** proposing?



What expertise does the group have?

What expertise does the group ***need***?

recruiting activities

Who do I know who has that expertise/authority?

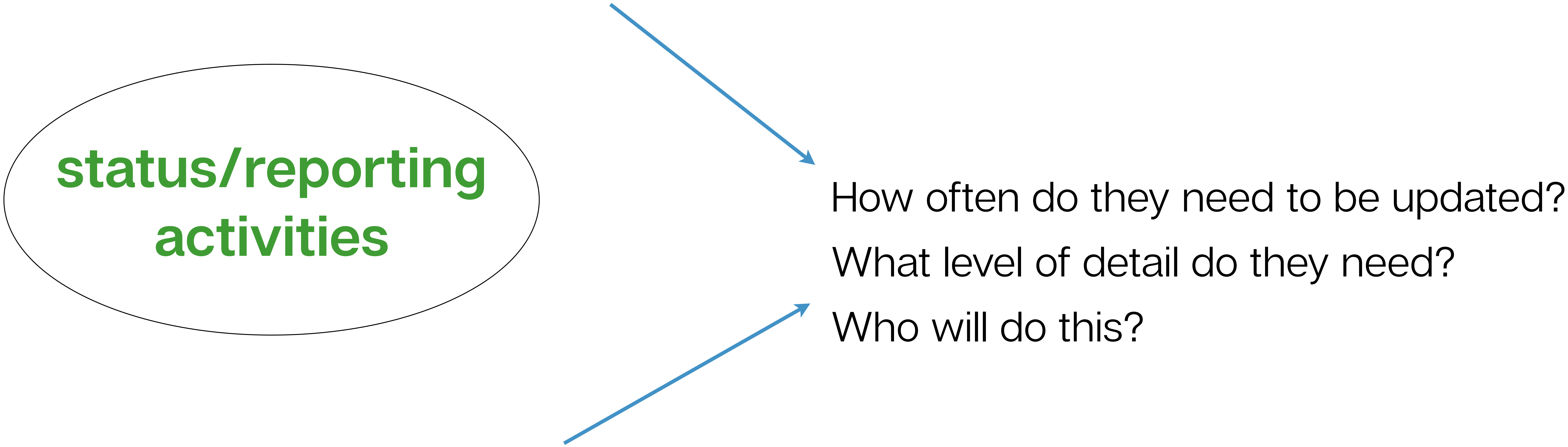
How can they be called on for help?

What do they need to know when they arrive?

What authority does the group have?

What authority does the group ***need***?

Which individuals or groups need to be informed about the ***current status of the response?***



status/reporting activities

How often do they need to be updated?

What level of detail do they need?

Who will do this?

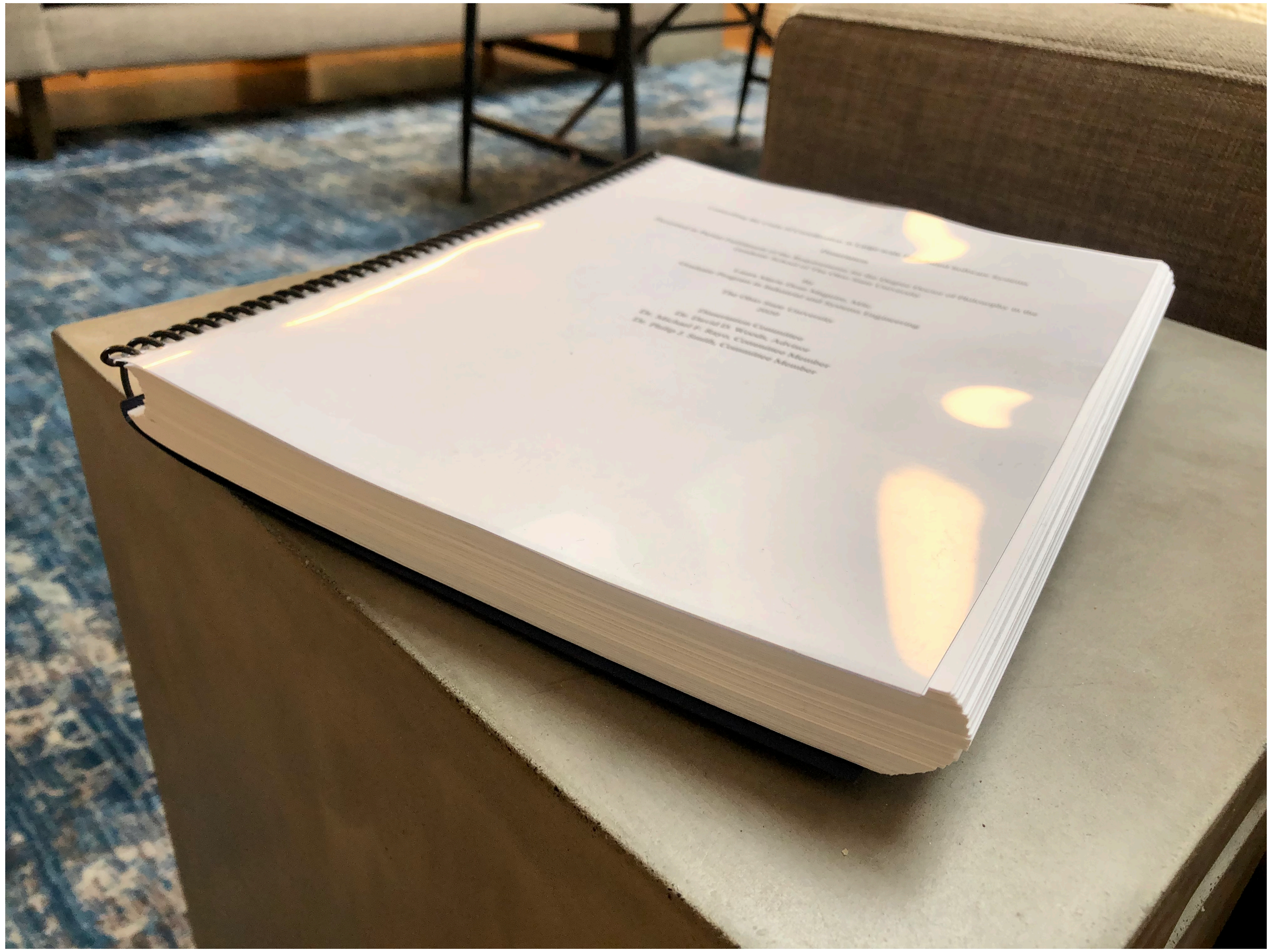
Which individuals or groups need to be informed about ***potential downstream impacts or effects?***

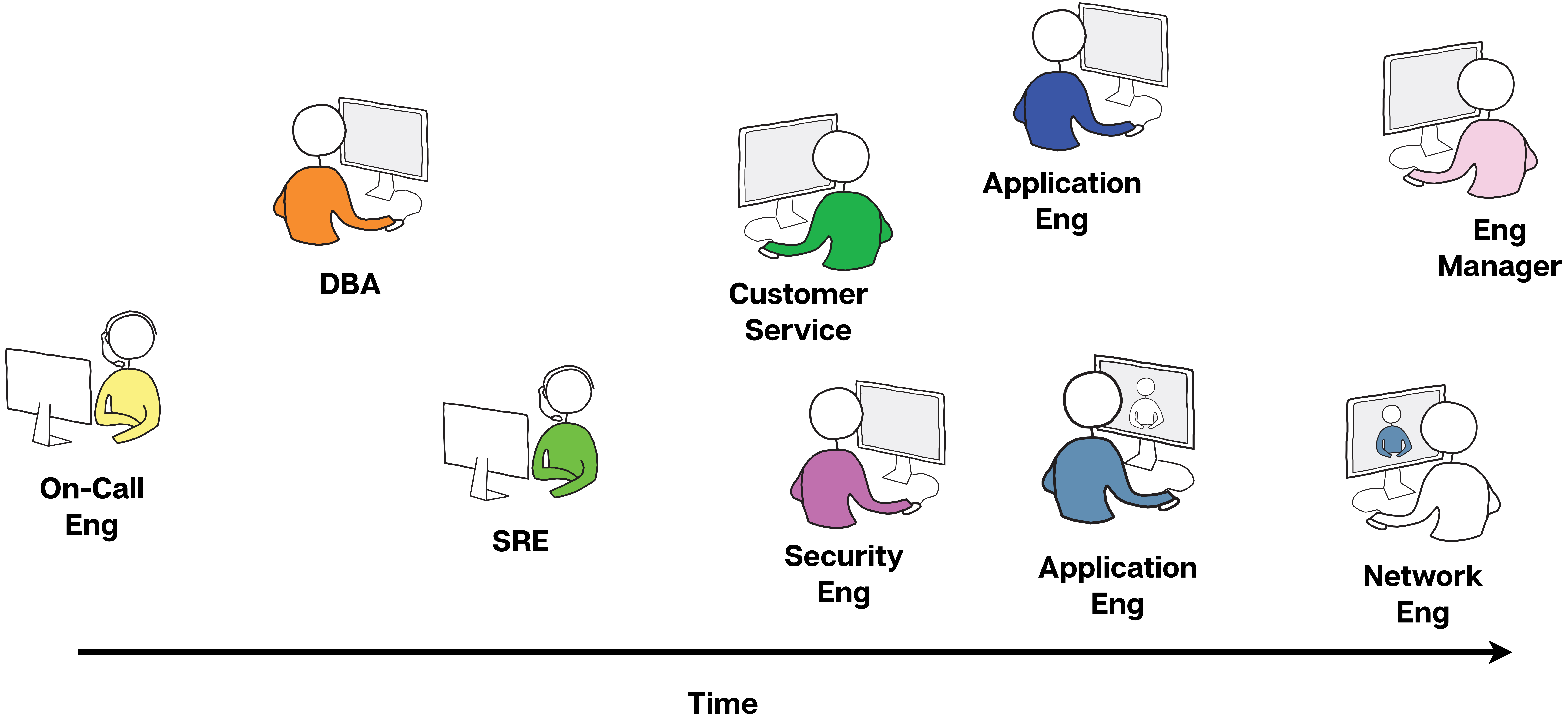
Costs of Coordination

Controlling the Costs of Coordination in Large-scale Distributed Software Systems

Dr. Laura Maguire

bit.ly/MaguirePhD





On-Call Eng

DBA

SRE

Customer Service

Security Eng

Application Eng

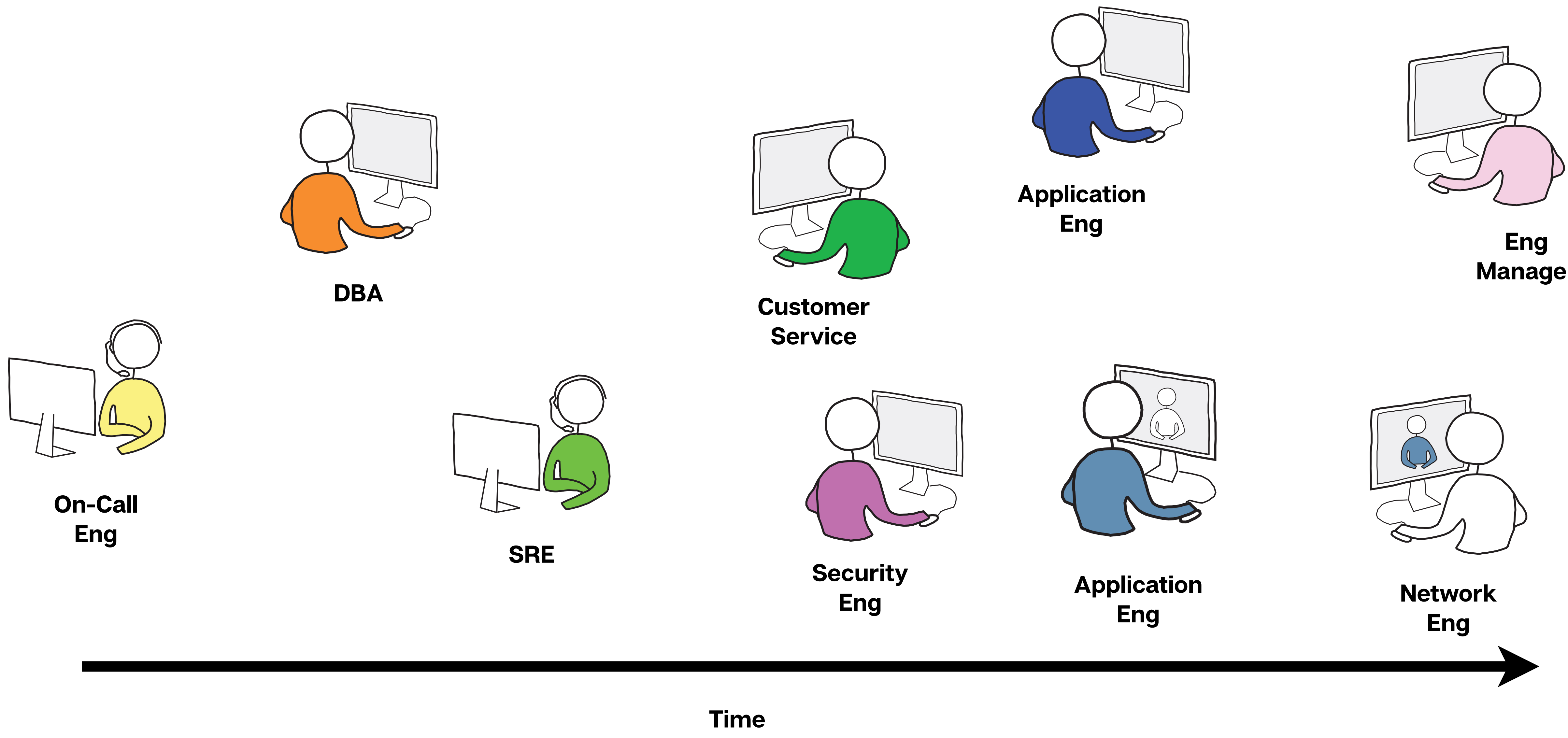
Application Eng

Eng Manager

Network Eng

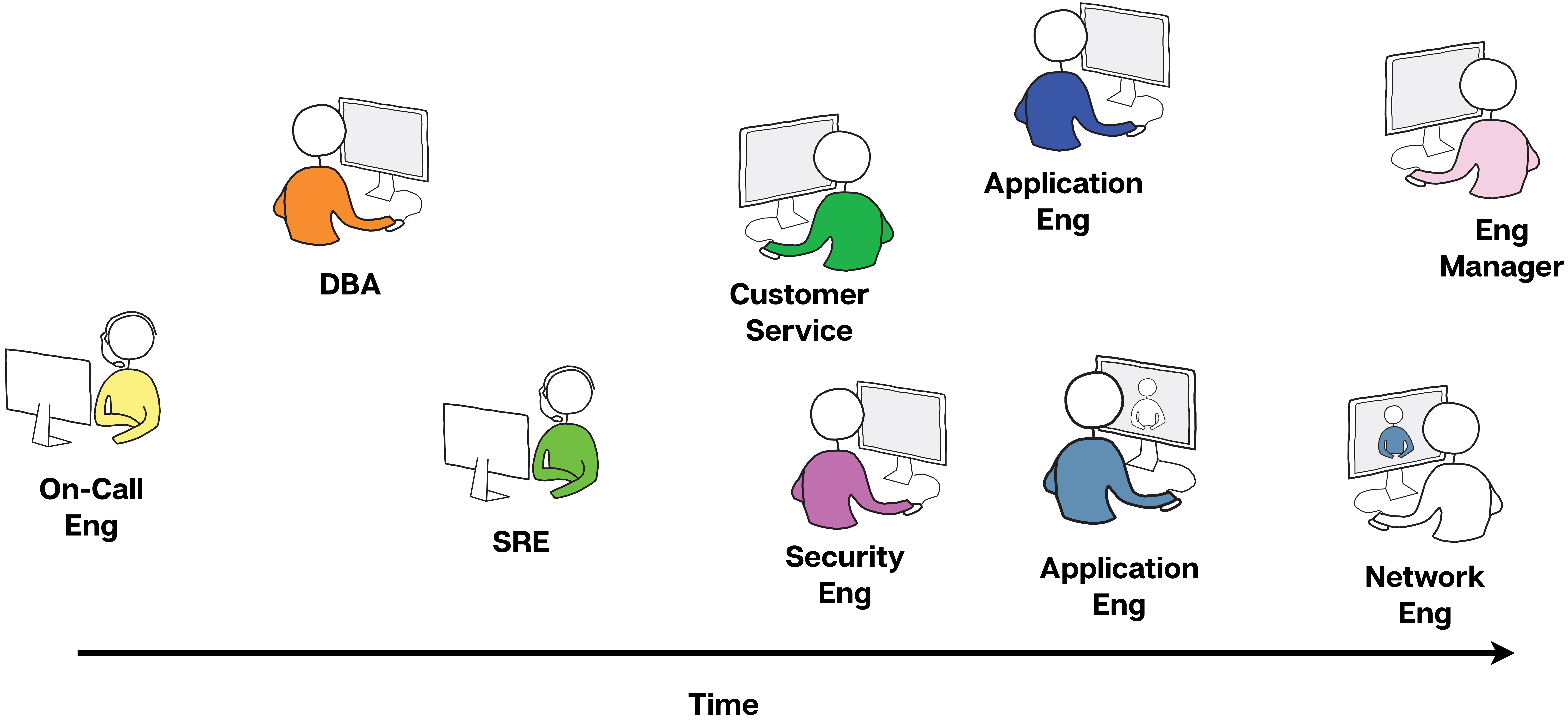
Time

effort needed to bring them "up to speed" ← **TRADE-OFF SPACE** → ... takes attention away from handling the incident

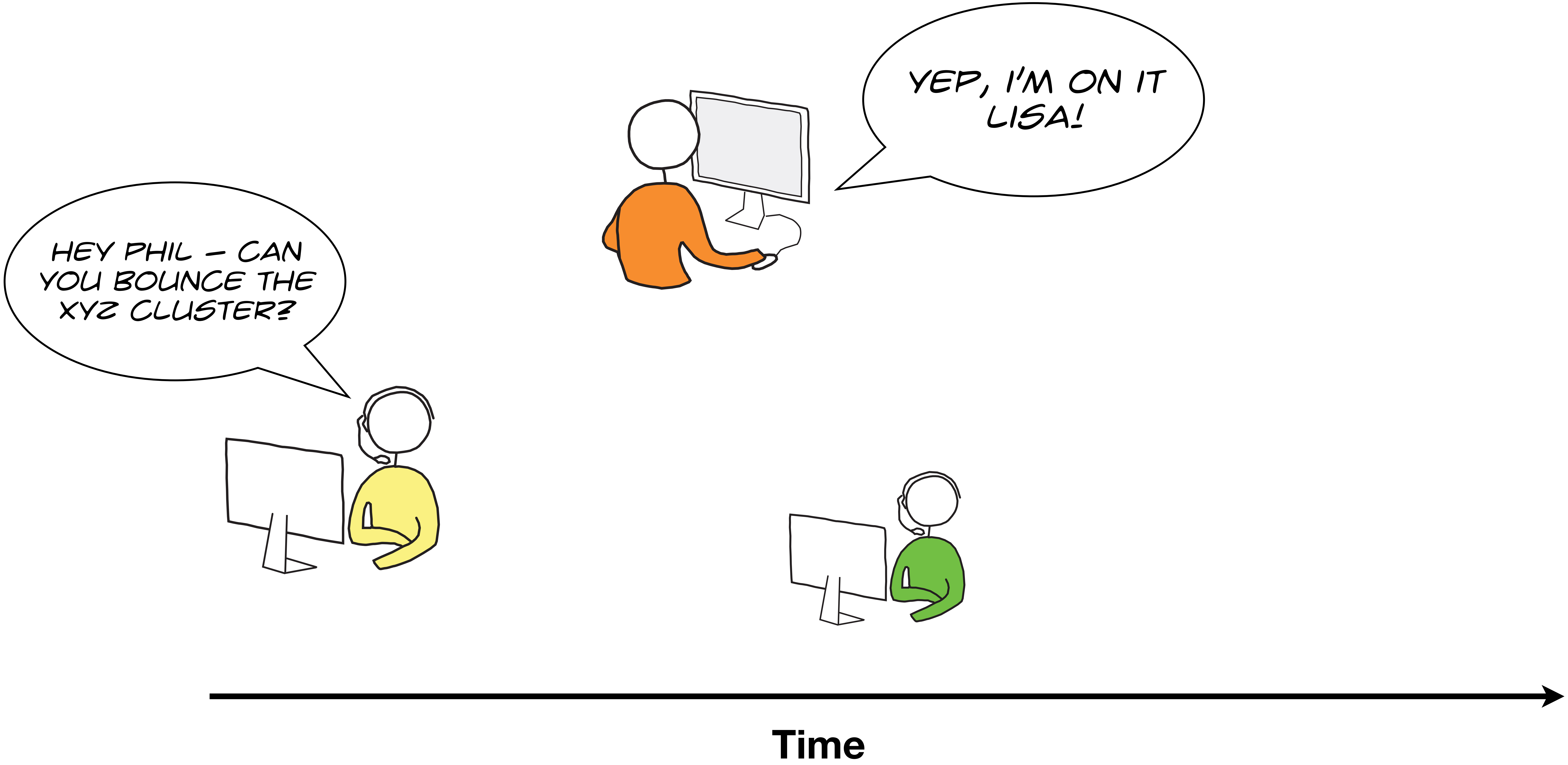


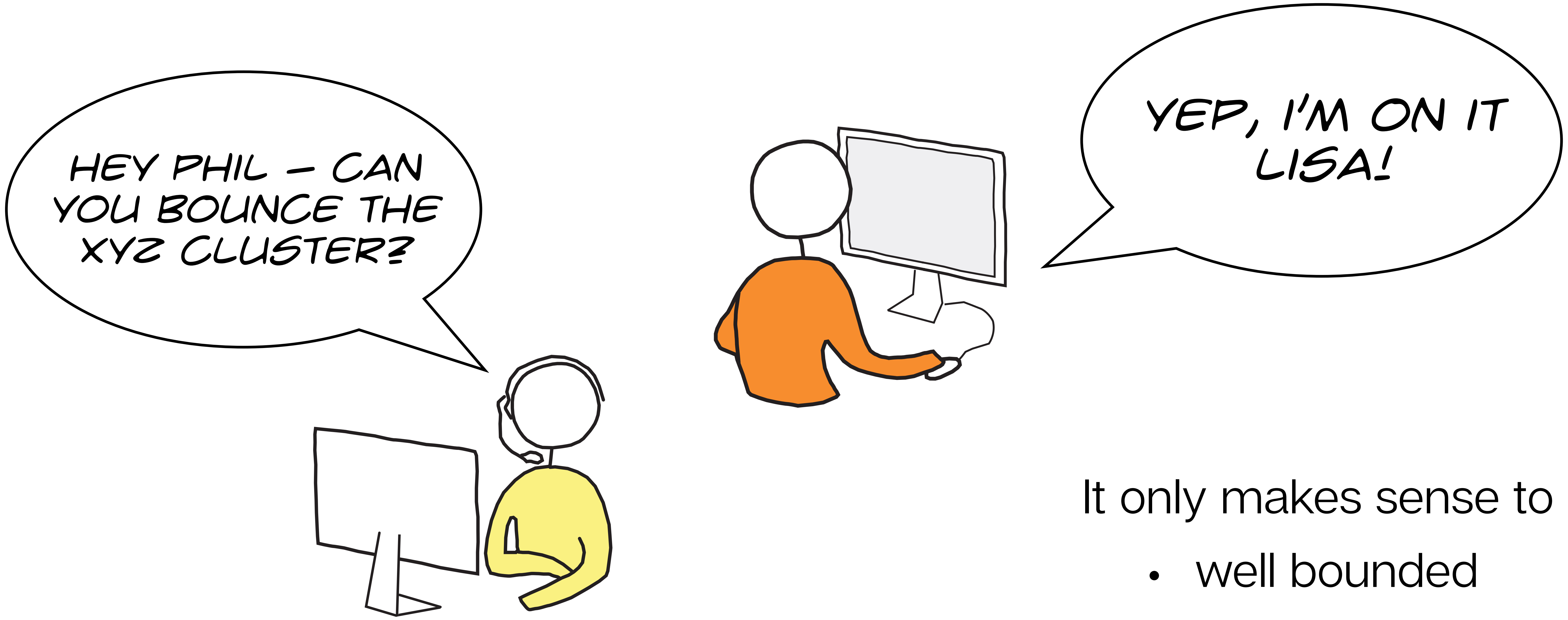
Should they stay focused on the incident in order to maximize their chances of quick diagnosis and repair….

…or devote some of their effort to bringing others up to speed so that they can help in that work?



“Divide and conquer” also has costs





Benefit:

Lisa can do other things while Phil works on that

Cost:

- have to identify the task to delegate
- have to select someone to do it
- have to specify what is to be done
- pay attention later to the report back from person

It only makes sense to assign tasks that are:

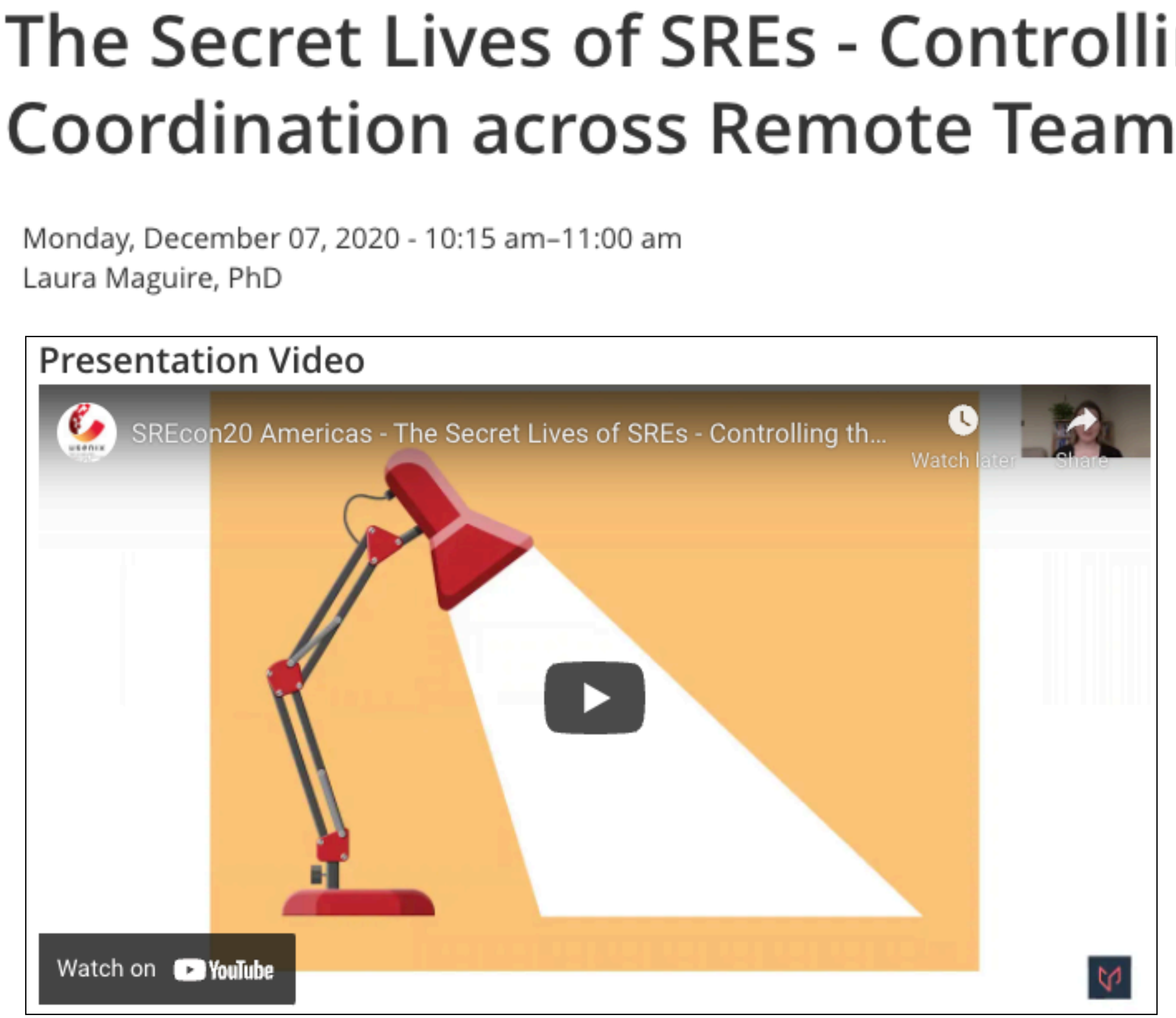
- well bounded
- can be accomplished by an individual, and
- for which a suitable person is both available and ***not already working on a higher priority task.***

There is yet another catch for this gamble:

New information about the event might reveal that doing a specific task could be unecessary...or even hazardous (!)

This imposes additional workload on all the parties.

Don’t take my word for it



The Secret Lives of SREs - Controlling the Costs of Coordination across Remote Teams

Monday, December 07, 2020 - 10:15 am–11:00 am
Laura Maguire, PhD

Presentation Video

SREcon20 Americas - The Secret Lives of SREs - Controlling th...

Sacrifice Decisions

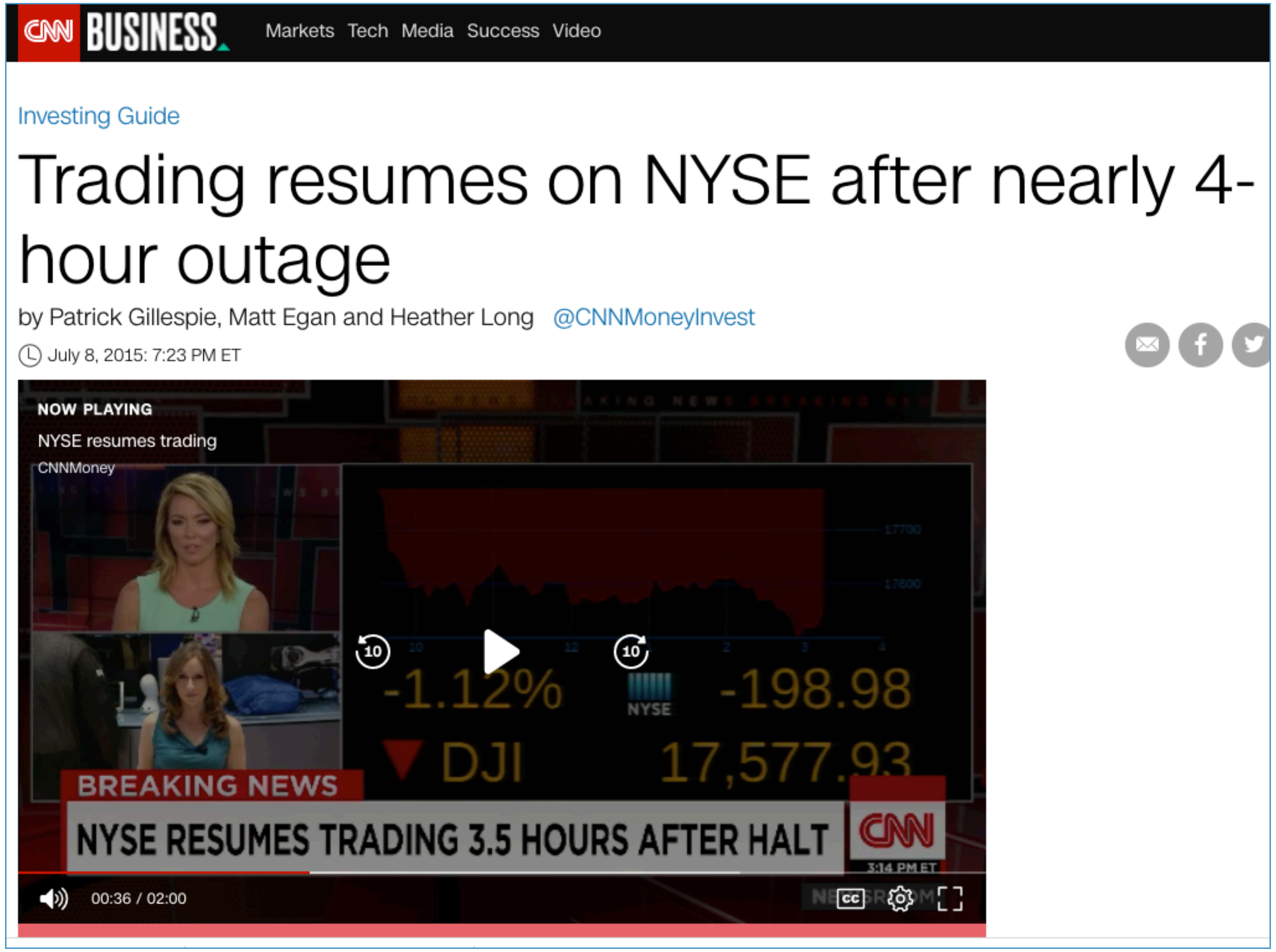
> “During disturbances…achieving important ("high level") goals may require abandoning less important ("low level") ones.
>
> Sometimes the sacrifice requires incurring damage, even severe damage, in order to prevent an even greater catastrophe.”

(Woods, D. D. (2017) *STELLA: Report from the SNAFUcatchers Workshop on Coping With Complexity*)

examples:

- Forcing a network partition to allow recovery
- Killing slow-running database queries until they can be fixed in code
- Reducing (or even eliminating) cross-datacenter encryption mechanisms temporarily to relieve data replication lag

Sacrifice Decisions



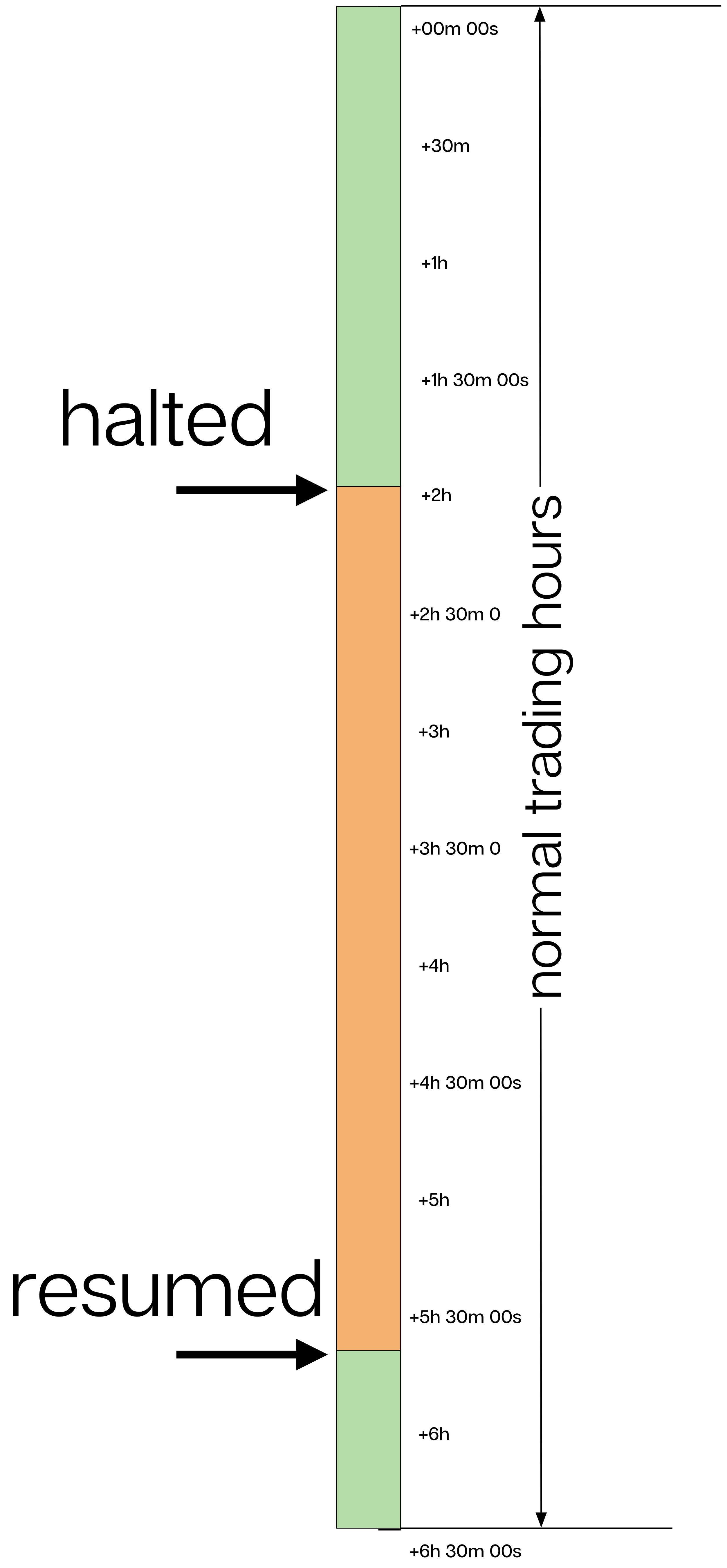
CNN BUSINESS Markets Tech Media Success Video

Investing Guide

Trading resumes on NYSE after nearly 4-hour outage

by Patrick Gillespie, Matt Egan and Heather Long @CNNMoneyInvest

July 8, 2015: 7:23 PM ET



halted

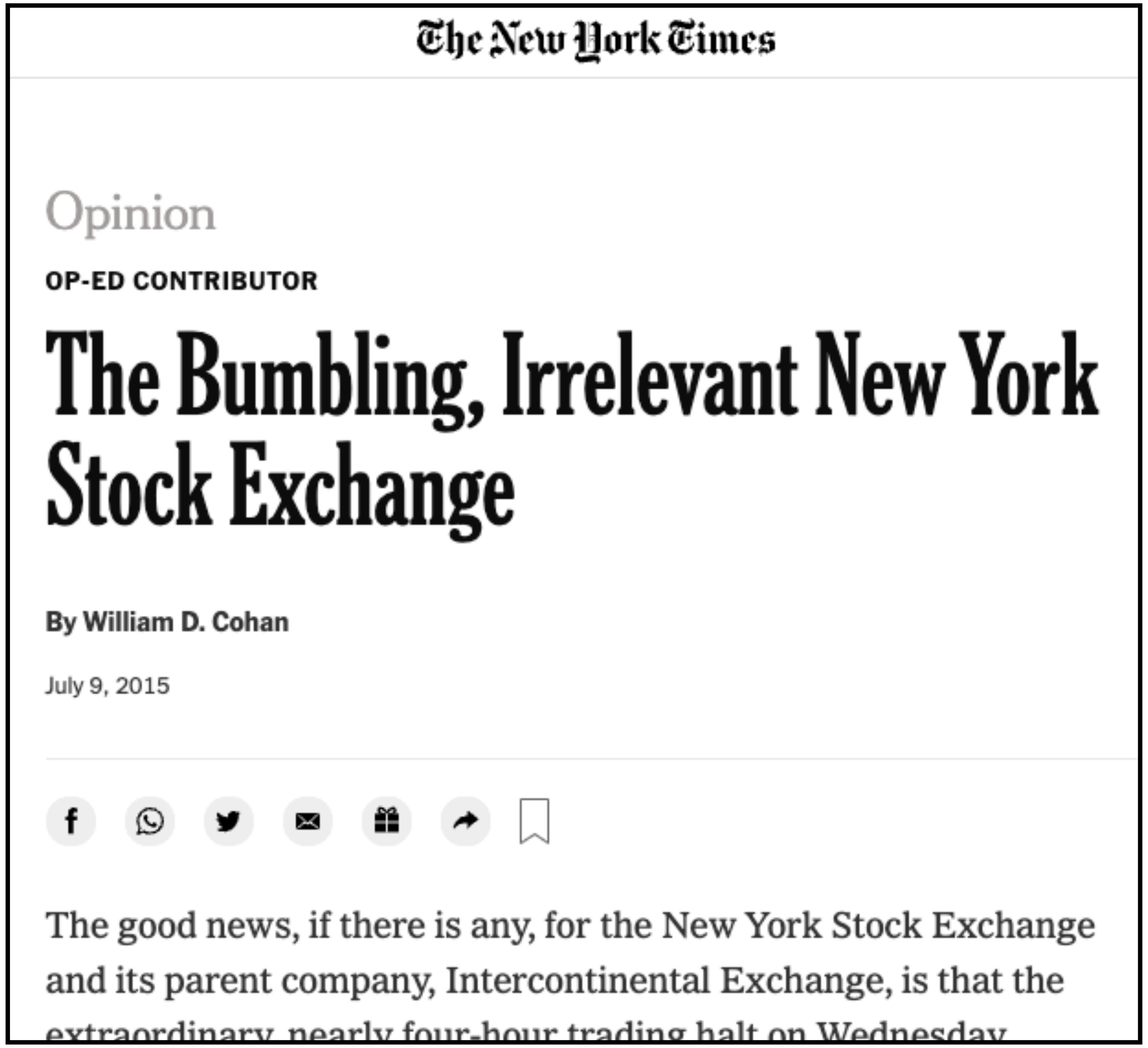
resumed

normal trading hours

Sacrifice Decisions

“My first concern was do no harm during the day," Farley said.

“Those stocks continue to trade elsewhere. Get the problem fixed. And get it back up and running for the close. We chose the least disruptive option for customers.”



The New York Times

Opinion

OP-ED CONTRIBUTOR

The Bumbling, Irrelevant New York Stock Exchange

By William D. Cohan

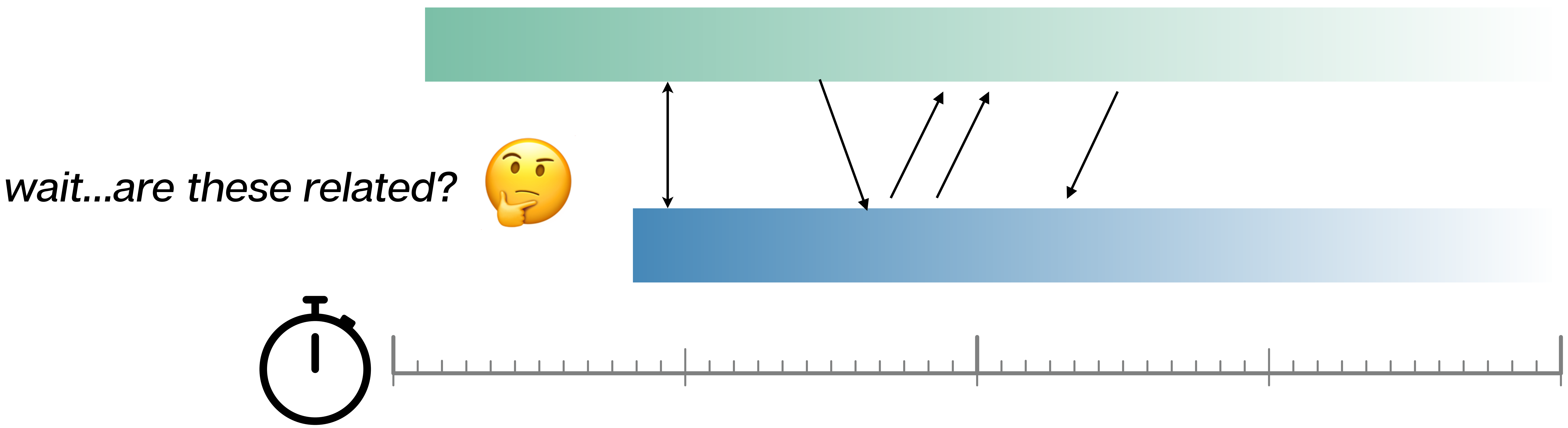
July 9, 2015

The good news, if there is any, for the New York Stock Exchange and its parent company, Intercontinental Exchange, is that the extraordinary, nearly four-hour trading halt on Wednesday

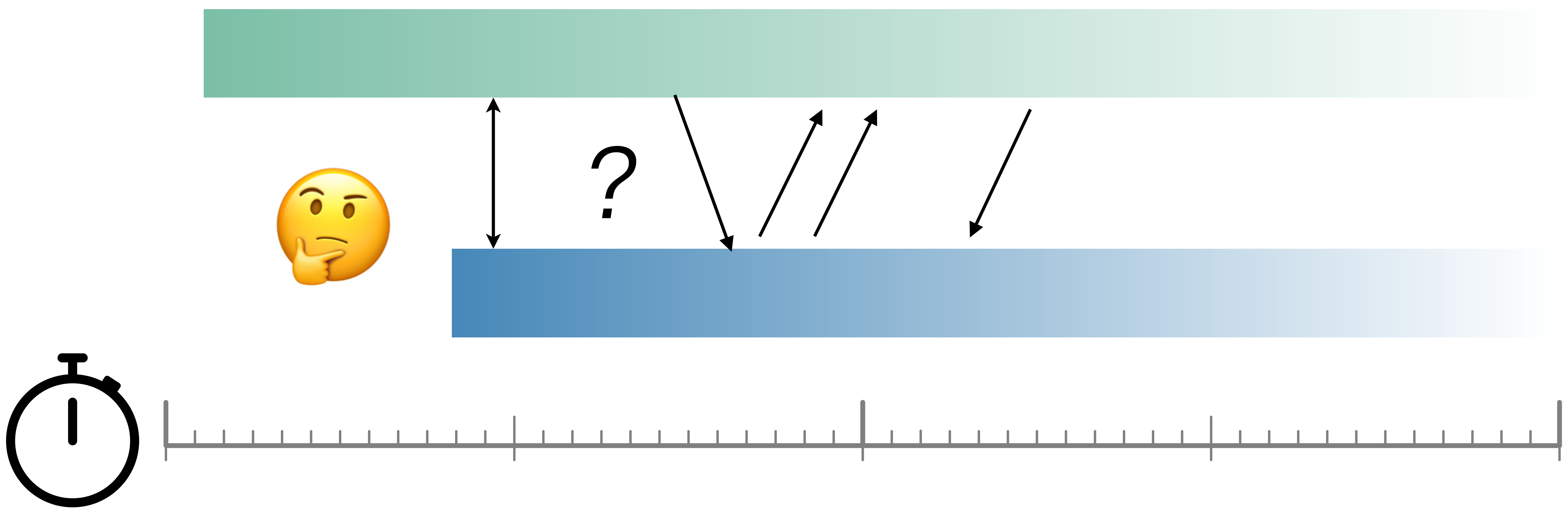
Parallel Incidents Dilemma

1. If two incident responses **are** related, combining efforts & observations could be *very* helpful and productive.

2. If two incident responses are **not** related, investigating if they were could be seen as a waste of time.



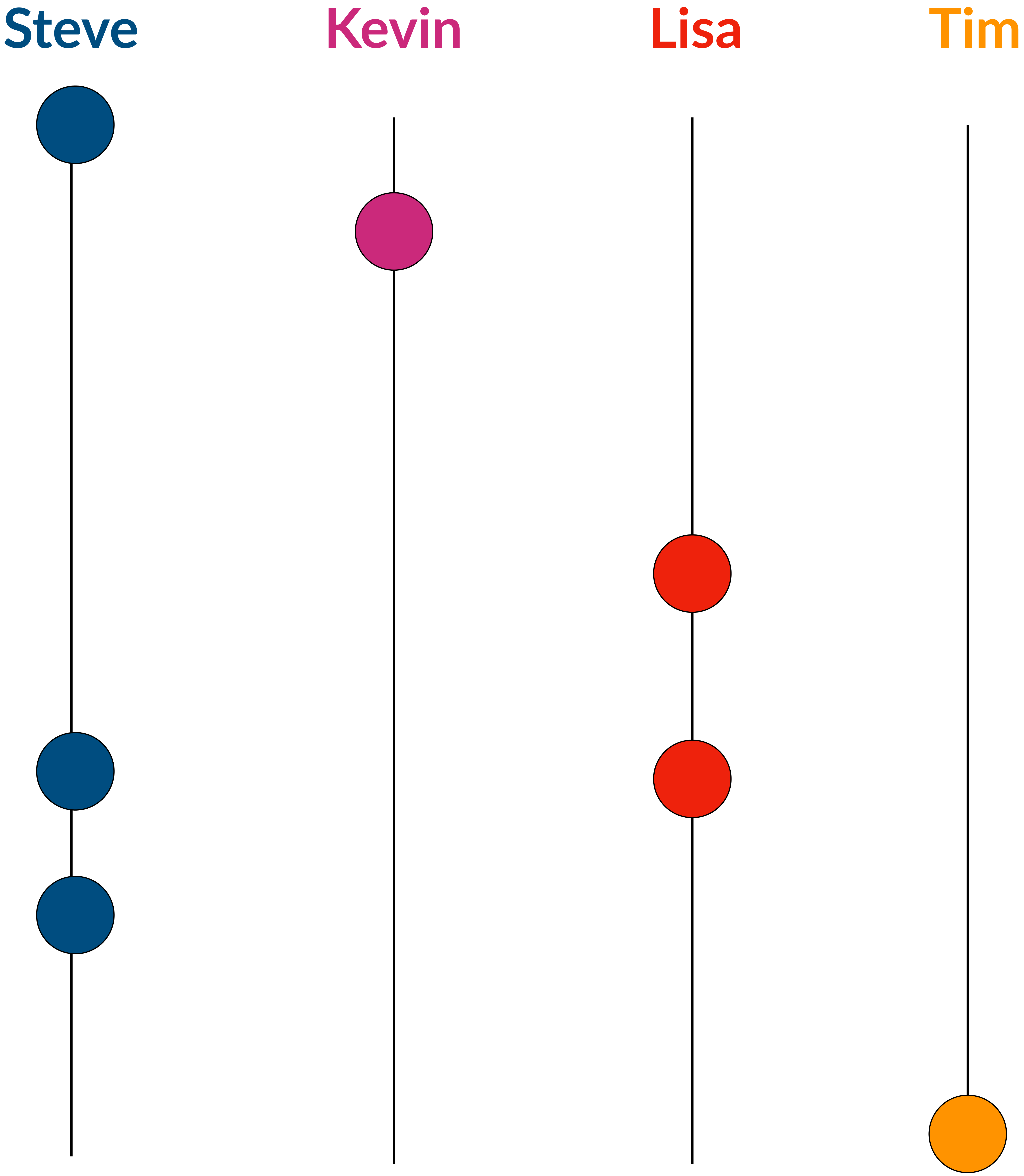
Parallel Incidents Dilemma

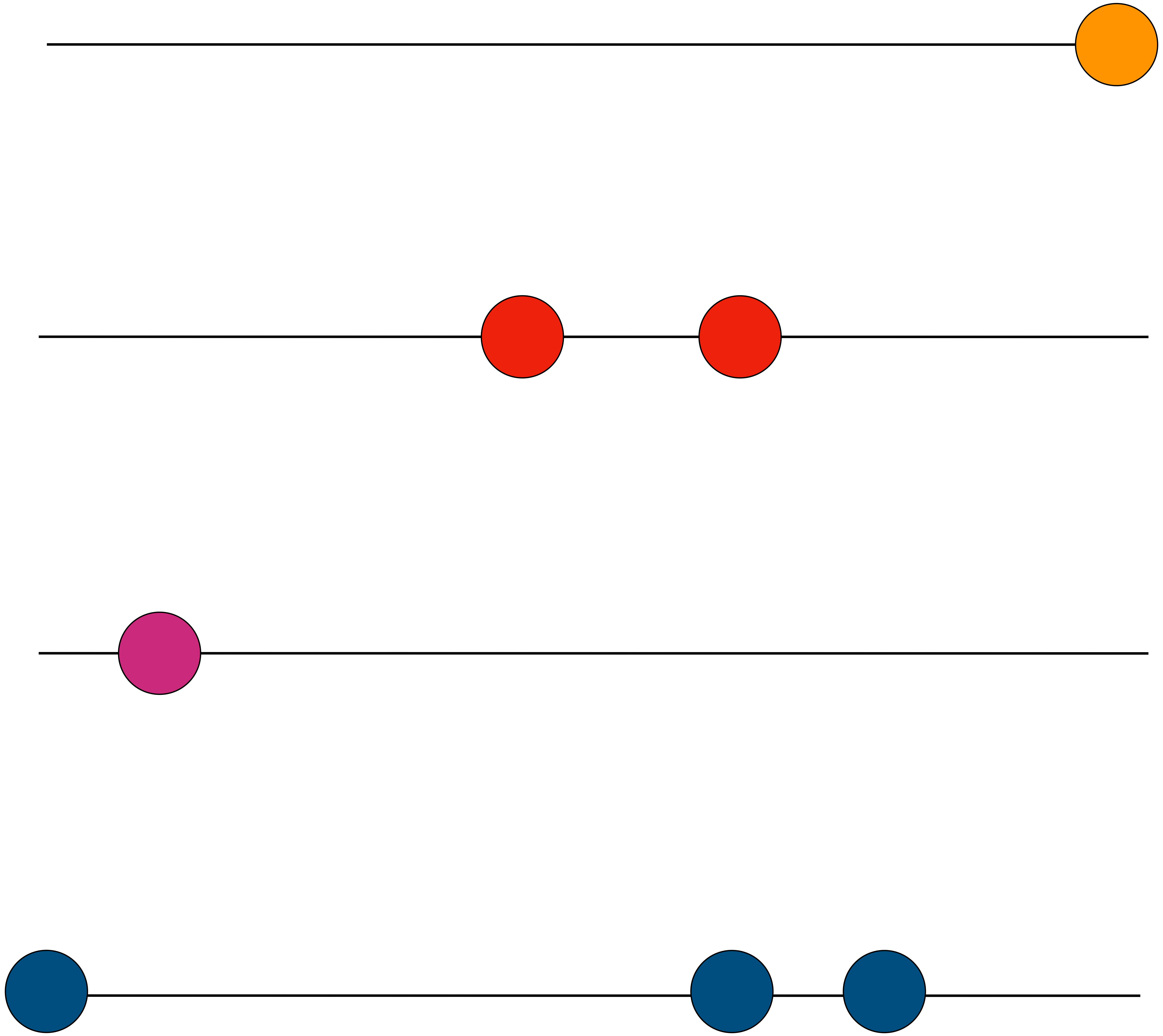


- How can you discover if *another* incident response is happening at the same time yours is?

- If you do discover one, how could you tell if time/effort spent determining if they are related is warranted?

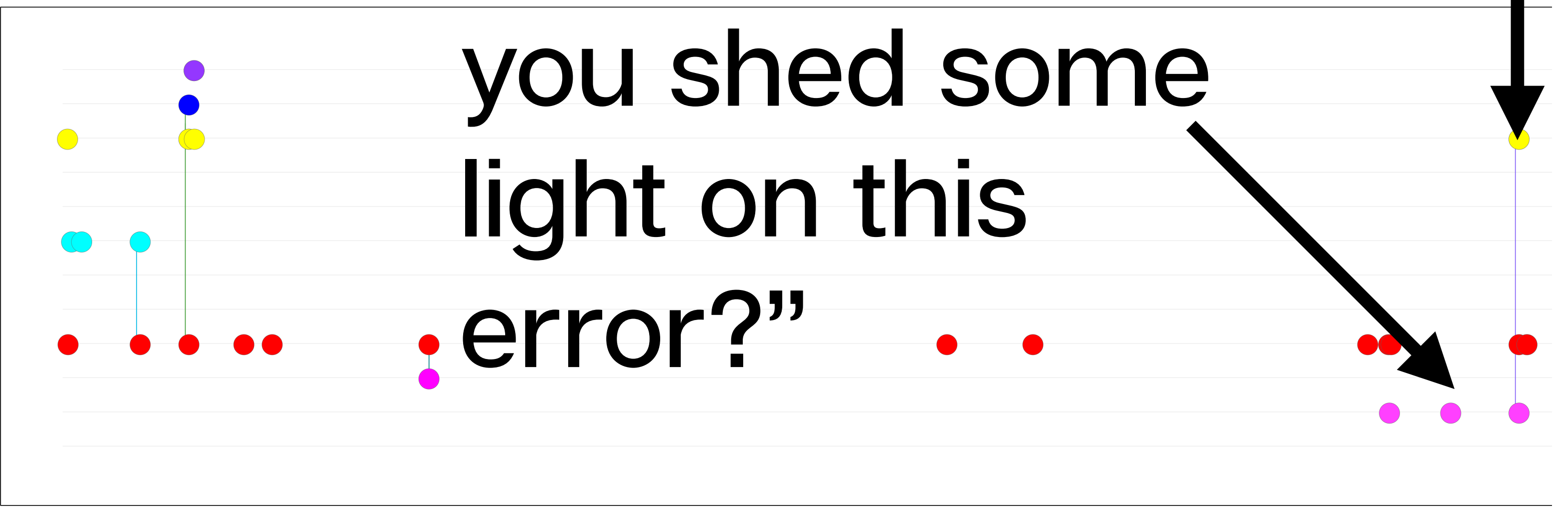
```
10:44:24   Steve   TOPIC :bunch of hosts flapping
10:44:38   Kevin   i checked my jobs; this is not the same outage like last Thursday
10:44:49   Lisa    all over the console on memcached21: nf_conntrack: table full, dropping packet
10:45:08   Steve   oh wow
10:45:18   Lisa    did anyone push anything iptables related?
10:45:20   Steve   is that recent though?
10:45:23   Tim     Lisa: I did
```


Steve
Kevin
Lisa
Tim

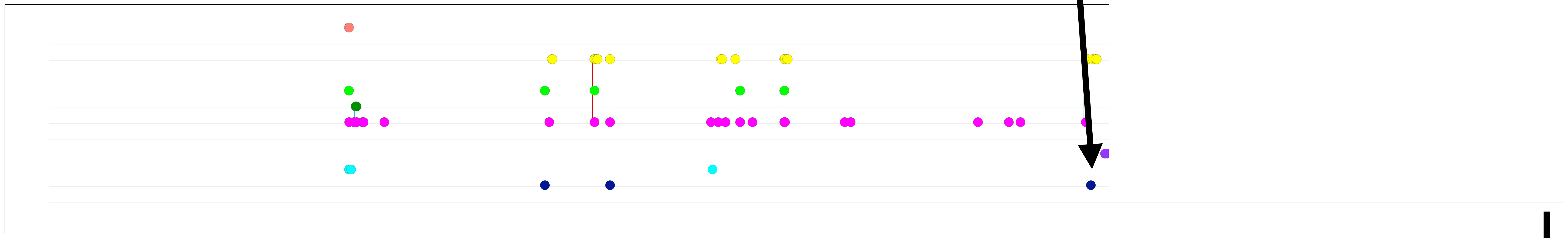


+01:20:54
“Hey Steve, can you shed some light on this error?”

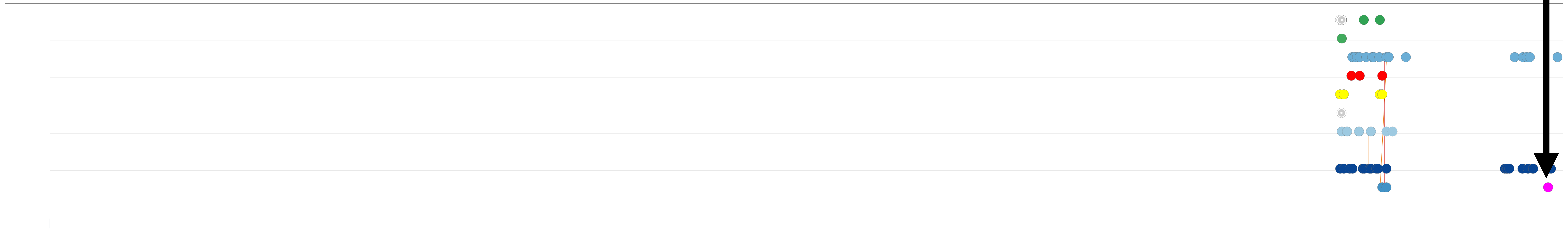
+01:24:42
“Ah, yeah there was a PR merged this morning…”



+02:20:26
“just joining so this might be off, but there’s another incident…”



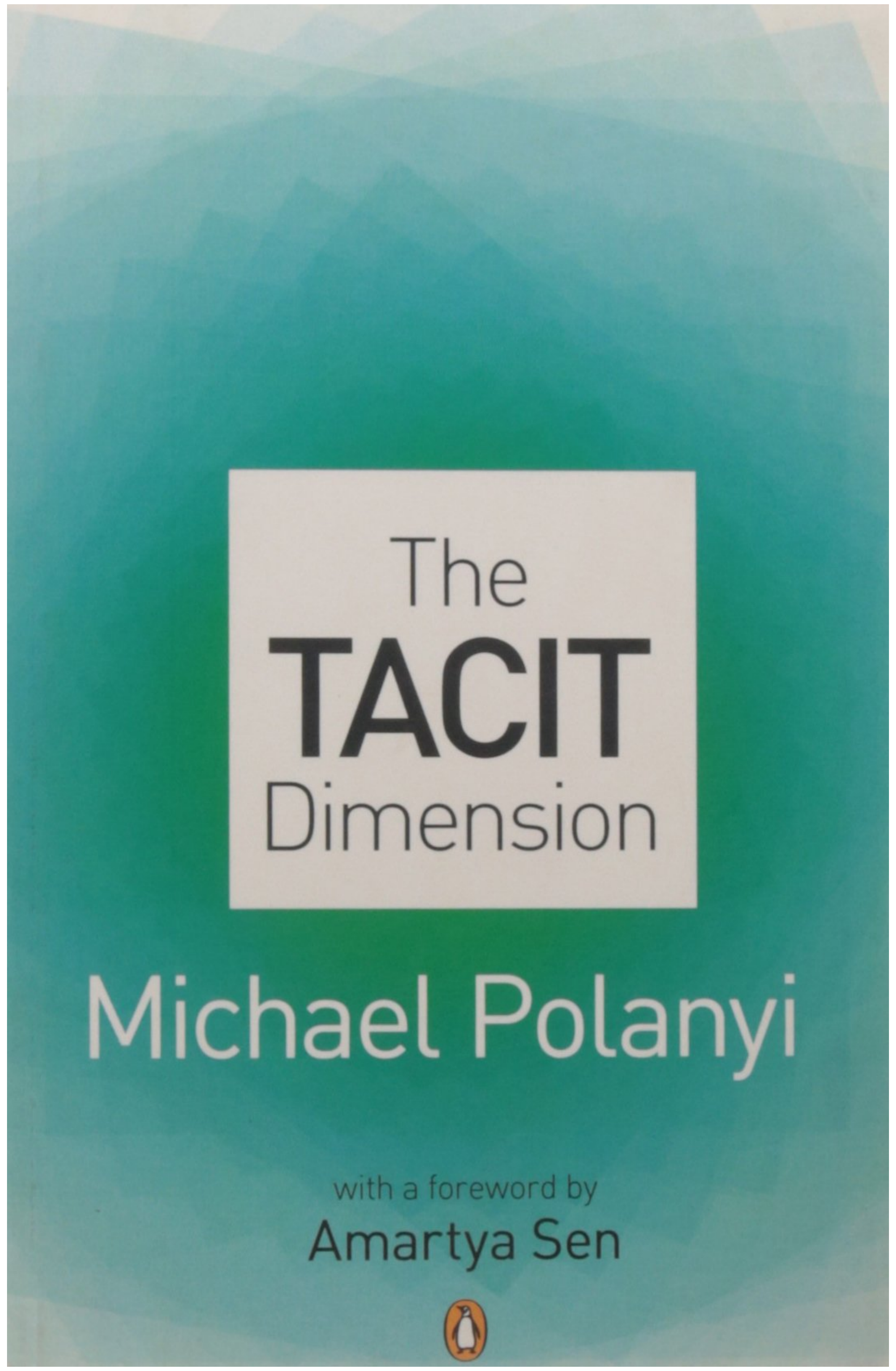
+03:16:13
“hey all…this is what’s happening…”



We are way better at this stuff than we think we are

It’s also difficult for us to see what makes us good at it

Expertise is more invisible than we realize.



“We can know more than we can tell.”

Having vocabulary for these phenomena is important.

When we’ve got words for them, we should use those in our stories.

Thanks.