



# Effort, Design & Responsibility

Dr Marcelo Thompson | HKU

20 Jan 2020

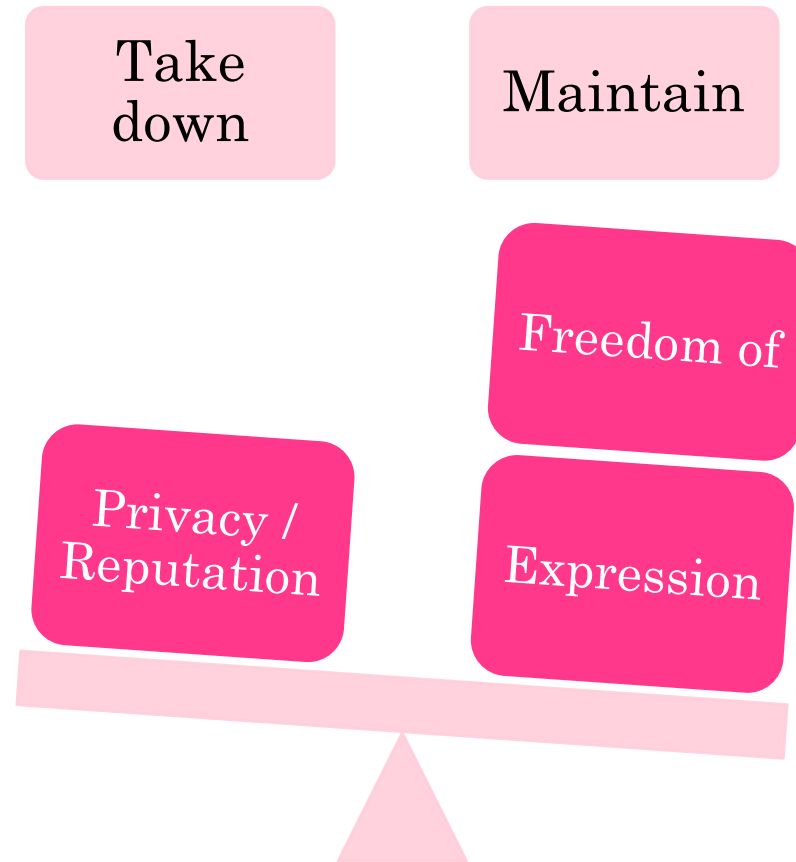
@KOREA UNIVERSITY



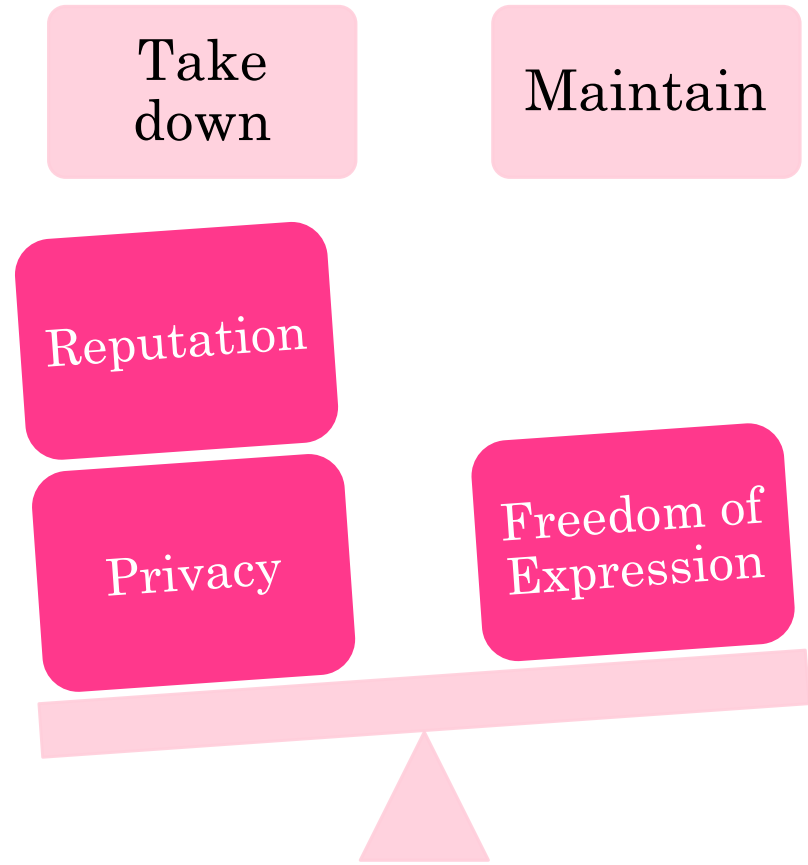
# 1. PLATFORM RESPONSIBILITY



# OUTCOMES-BASED LIABILITY REGIMES





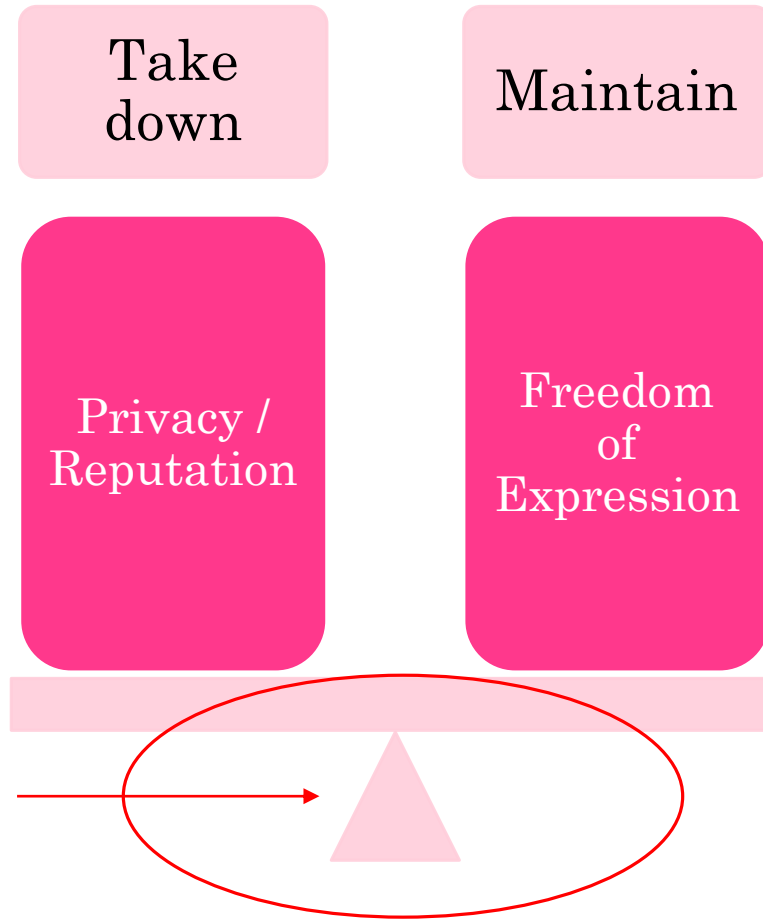






# **EFFORTS-BASED LIABILITY REGIME**







# Reasonable Efforts Approach



# Cushioning System for Honest Mistakes

# Wider Possibilities for Sanctioning

E.g. Fines, Warnings,  
Incentives



**Duty of Care: Founded on a normative idea of negligence – and on a commitment of normative integrity**

*“[a]n approach that ... transcends the factual dimension of content takedown to focus on the matter of how an Internet intermediary lives up to its normative commitments: how rightly or wrongly it traverses the normative space of [its] duty of care”.*

*“The normative negligence approach recognizes the difficulties in interpreting the facts (the nature of content) in light of the normative order (its illegality or illegality). Yet, while paying heed to this difficulty, this approach does not refrain from requiring a commitment of normative integrity”.*

*“...a pursuit of coherence between the intentional dimension of the artifacts intermediaries design and the broader set of reasons that compose the normative order”*

*... a commitment of trying assiduously enough to succeed in understanding and evaluating the facts brought before them, in coherence with the central normative commitments of the communities they inhabit”.*





**Scalability: i.e. Calibration in  
Accordance with Platform's  
Econ / Techno Possibilities**

**AI could be a factor**



## **2. RECENT DEVELOPMENTS**

## GERMANY: NetzDG 2017

- Provider of social network (with more than 2 million users in Germany) must maintain an **effective and transparent procedure for handling complaints** about unlawful content
- Handling of complaints to be monitored via **monthly checks** by the social network's management.
- Organizational deficiencies in dealing with incoming complaints must be immediately rectified.
- Training courses and support programmes for moderators on a regular basis
- Regulatory **offence** deemed to have been committed by any person who, intentionally or negligently ... **fails to rectify an organisational deficiency** or to rectify it in due time

## UK: ONLINE HARMS WHITE PAPER (1)

- New system of accountability and oversight for tech companies, moving beyond self-regulation
- Make clear companies' responsibilities to keep UK users safer online, with most robust action to counter illegal content and activity
- Independent regulator to set clear safety standards (codes of practice), reporting requirements (transparency reports), enforcement powers
- First attempt globally to address a comprehensive spectrum of online harms in a single and coherent way
- Coherent, proportionate and effective approach
- Boost technology sector / technology as part of the solution

## UK: ONLINE HARMS WHITE PAPER (2)

- Responsible digital design
- Culture of continuous improvement / private-run platforms akin to public spaces
- New statutory duty of care to make companies take more responsibility for the safety of their users and tackle harm
- Risk-based and proportionate approach, with initial focus on companies that pose the biggest and clearest risk -- though every company will need to fulfill its duty of care
- Expectation of doing what is reasonably practicable to counter harmful activity or content, depending on the nature of the harm, the risk of the harm occurring on their services, and the resources and technology available to them



# 3. INTERMEDIARY AS DESIGNERS

# INTERMEDIARIES AS DESIGNERS

- Fading distinction between a company pursuing its **own ends** and **merely acting as an intermediary**
- Platforms as **medium** and **message** of our social and political processes  
E.g. fake news, political advertising
- Content moderation as a matter of design

## AI AS THIRD-PARTY?

- In a **strong AI-based world**: **difficulties** for intermediaries/designers in avoiding responsibility for **third-party content**
- **AI itself** as a third party?
- Real question: when things go wrong, have they gone wrong due to **external circumstances, in spite of** intermediaries / designers best efforts?
- Can intermediaries/designers point to technological processes and say: I couldn't help it; **that was not me**?





# 4. CHALLENGES

## EFFORT & FORESEEABILITY

- Literature emphasizes how liability in law has been based on an assumption of **foreseeability** of future events
- Ensuing idea that it is **reasonable to guard against** these events
- But **effort** is also a foundation for liability
- Negligence in law is **a failure to try assiduously enough to avert** (limit, reduce, control) the **unwelcome side-effects of one's** (otherwise valuable) **endeavours**" (Gardner, 2005).
- What are the boundaries between effort and foreseeability?

## VARIABLE STANDARDS

- Tort law excludes **circumstances internal to an agent** (e.g. mental incompetence) from consideration by courts in excluding liability
- Exception: **children** – “limitations ... as being characteristic of humanity at his stage of development and in that sense normal” (*McHale v Watson*).
- Should **different agents different limitations** in relation to their command of technology **count differently** (are they **internal**)?
- Or **will AI make us all into children**?



**Thank you**

Marcelo Thompson | HKU

20 Jan 2020