

UNDERSTANDING THE SOCIAL ATTACK SURFACE: NORMS AND CRITICAL SOCIAL INSTITUTIONS

John Symons

Director, Center for Cyber Social Dynamics

johnsymons@ku.edu

University of Kansas



1912



1920



1923



1929



1941



1946



2006-now

TOPLINE ARGUMENT

DEFENSE OF THE UNITED STATES REQUIRES A SCIENTIFIC UNDERSTANDING OF INTERVENTIONS AGAINST SOCIAL INFRASTRUCTURE (ESPECIALLY THE CYBER-SOCIAL INTERFACE)

- Social infrastructure includes critical social institutions, norms, and choice architectures
- The social attack surface has been understood in individualistic terms (social hacks/disinfo & misinfo). However, social norms and consequently institutions are also targets for our adversaries.
 - via social media and traditional media
 - machine learning / marketing campaigns / platform design
- Defensive strategies include tools for tracking campaigns against social norms, forecasting norm change, automated responses, understanding and protecting vulnerable individuals and communities through controls on platform design.

RECOGNIZING THE ROLE OF THE SOCIAL

- **Western focus has been on espionage rather than warfare** (Russia and China see this very differently)
 - Current approaches to the “human factor” in security science focus on individual psychology (individual epistemic states, opinions, emotions, attitudes) and individual decisions while neglecting the importance of social norms.
- Social infrastructure depends on social norms
 - We study efforts to modify norms and we attempt to identify vulnerabilities

THE SOCIAL ATTACK SURFACE.

Given a sufficiently long, timeframe social infrastructure is more important to security than physical infrastructure

and

protecting it requires that we understand the norms, expectations, and choice architectures (especially at the cyber-social interface) that constitute social institutions.

THEORETICAL FRAMEWORK

INITIAL FINDINGS

COUNTERINTUITIVE RESULTS

- Institutions are animated by **social norms**. (Symons and Elmer 2023, Bichierri 2018)
- Evidence that there are coordinated Russian campaigns against critical institutional norms (initial analysis of Twitter data from known Russian associated accounts)
- Tools for forecasting norm change (Observatorium)
- Studying *live* norms requires indirect strategies – they play a role in explanations of behavior
 - Why did she buy that white handbag? Because she’s getting married.
- One way to undermine a norm is to defend it
 - “No public defecation in the philosophy department please”
 - “Back the Blue”

NORMS AND INSTITUTIONS SYMONS AND ELMER (2023)

- Institutions are animated by **social norms**.
- Without the relevant constitutive social norms in place, legal and institutional regulation will not suffice for most of the institutions we care about.
 - Social equilibria develop to **support** “rule of law” or “compliance”
- Not by convention: Failure to comply matters normatively whereas breaking mere conventions is not blameworthy in most cases.
- Not moral: Even if people judge some set of actions (say bribery) to be immoral, the moral judgment can be subordinated by the empirical expectation
 - “Everyone is doing it, I know it’s wrong, but what can you do? If you want to get your permit you have to give a little bribe to the officials”

WESTERN LITERATURE ON THE SOCIAL ATTACK SURFACE TO DATE – INADEQUATE BUT USEFUL

“The **social engineering attack surface** is the totality of an individual or a staff’s vulnerability to trickery. **Social engineering attacks** usually take advantage of human psychology: the desire for something free, the susceptibility to distraction, or the desire to be liked or to be helpful.” (Edwards et. al 2017)

Attacks on Google (2009), RSA (2011), JP Morgan, (2014) Ukrainian power grid (2015) , etc., all leveraged social engineering hacks

Email: phishing/spear-phishing emails

Telephone: voice phishing

Physical: gaining physical access through use of a deceptive pretext, or via physical media

Western literature has focused on individual psychology and vulnerability and has understood these questions within the framework of cyber-espionage.

COMMON KNOWLEDGE, NORMS, AND SOCIAL LIFE

Many prominent features of social life depend on agents assuming that the rules of some practice or norm are matters of *common knowledge*.

Drivers know that a red traffic light indicates that they should stop at an intersection. However, for the convention of traffic lights to be in place at all, it is first necessary that drivers must also know that other drivers know that red means stop . In addition, drivers must also **know that everyone knows that everyone knows that...**

The conventional role of traffic lights relies on all drivers knowing that all drivers know the rule, that the rule is a piece of *common knowledge*.

A variety of norms, social and linguistic practices, agent interactions and games presuppose common knowledge.

The formal strategy for thinking about common knowledge is quite different from what is employed in network science.

BICHIERRI'S DEF (2015)

- “A social norm is a behavioral rule ‘R’ that applies to a certain social context ‘C’ for a given population ‘P’. People in the population prefer to follow the rule in the appropriate context if they believe that a sufficiently large part of the population follows the rule (**empirical expectations**), and further, if they believe that other people think that they ought to follow the rule, and may sanction them if they don’t (**normative expectations**).”
- For example, in Western countries, brides traditionally wear white at weddings because nearly all other brides do (empirical expectations) and they believe that others think that they should wear white (normative expectations).

WHAT ARE SOCIAL NORMS FOR? SOCIAL NORMS CAN SOLVE SOCIAL DILEMMAS.

- They can resolve the tension between what is the best choice for the group and the best choice for the individual. By aligning our incentives, they make it possible for us to coordinate on the best outcome.
- **Social norms compose equilibrium states for human behavior**
 - which means that they are sustainable in the long run. Once we reach equilibrium, it becomes difficult to leave it.
- “A social norm **might be enforced** by informal social sanctions that range from gossip to open censure, ostracism and/ or dishonor.” Social sanction can motivate individuals to follow a norm out of fear of punishment or out of a desire to please and thus be rewarded.
 - Typically live norms need little enforcement
- An equilibrium is a situation that involves several individuals or groups, in which each one’s action is a best reply to everyone else’s action. It is a situation of stable mutual adjustment: Everyone anticipates everyone else’s behavior, and all these anticipations turn out to be correct.

WHEN DO PEOPLE FOLLOW SOCIAL NORMS?

if they believe that a sufficiently large part of the **relevant** population follows the rule (empirical expectations)

and further,

if they believe that other people think that they ought to follow the rule (normative expectations).

In these contexts the norm is resilient (LIVE)

ATTACKING THE NORM INVOLVES

Undermining empirical expectation:

We no longer believe that a sufficiently large part of the population follows the rule

and further,

Undermining normative expectations:

We no longer believe that other people think that we ought to follow the rule

ATTACKS ON NORMS CAN BE STUDIED

Detect attempts to undermine empirical expectations that support norms:

Goal of the intervention:

We no longer believe that a sufficiently large part of the population follows the rule

Detect efforts to undermine normative expectations:

Goal of the intervention:

We no longer believe that other people think that we ought to follow the rule

ATTACKING SOCIAL NORMS

- undermining expectations - changing the empirical and normative narrative
- creating the impression of contentiousness – the norm is no longer live
- identifying trendsetters - finding the relevant agents/communities
 - achieving a critical mass of norm violation
 - Some number of violations are always tolerable

LINES OF CURRENT RESEARCH

- How do social norms change?
 - Interdisciplinary review
- Can we track and predict changing social norms?
 - Observatorium Data – Immigration debates in Europe
- Can we identify adversarial efforts to intervene in social norms?
 - Social Media – Defund the police

IDENTIFY CORE INSTITUTIONAL VALUES AND MISSION – WHAT ARE THE RELATED NORMS?

Defend:

1. Discourage restrictive on choice architectures in cyber-social interface (false dilemmas and ignorant institutional design)
2. Identify “trendsetters” and provide greater reasons to identify with the institution and social norms
3. Respond to direct attacks on core institutional values

Attack:

Restrict choice architectures in cyber-social interface

Change view of expectations and perception of incentive structures (“only suckers do...” “everyone knows... is corrupt”)

Encourage trendsetters to violate, increase perception of options/possibilities (diminish their sense of stake in the existing norms)

UNDERSTANDING RELEVANT VARIATIONS IN THE TARGET POPULATION

- Variation in a population with respect to expectations
- Variation in a population with respect to commitment/stake
- Variation in a population with respect to connectivity/influence

TRENDSETTERS/**NORM** **VIOLATORS**/CONSERVATIVE LEADERS

- Variation in a population with respect to expectations –low expectations (?)
- Variation in a population with respect to commitment/stake –low commitment
- Variation in a population with respect to connectivity/influence – low influence

TRENDSETTERS/NORM VIOLATORS/CONSERVATIVE LEADERS

- Variation in a population with respect to expectations – high expectations
- Variation in a population with respect to commitment/stake – high stake
- Variation in a population with respect to connectivity/influence – high influence

TRENDSETTERS/NORM VIOLATORS/CONSERVATIVE LEADERS

- Variation in a population with respect to expectations –low expectations (?)
- Variation in a population with respect to commitment/stake –low commitment
- Variation in a population with respect to connectivity/influence –**high connectivity**

TRENDSETTERS/NORM VIOLATORS/CONSERVATIVE LEADERS

Trendsetters are the interesting mixed case:

Agents who have **false expectations** for some reason

Idealists, young people, protected members of a group

Cynics

Low stake, e.g rich enough to have walk-away money

Popular enough not to worry about social isolation or indifferent to social affiliation

Unusually courageous

With **high degree of connectivity or influence**

DEFENDING THE NORM

- Identify vulnerables and minimize the damage they can do
- Avoid cyber-social interface architectures that corrode valuable norms
 - apps and processes that encourage lying for example
 - security theater, covid theater, etc.
 - choice architectures in social media that promote concern with the views of one's affiliation group

RUSSIAN/SOVIET INFORMATION WARFARE

THE STORY OF 'THE TRUST' DEPLOYED BY THE CHEKISTS AGAINST WHITE RUSSIAN ÉMIGRÉ OPPOSITION IN THE 1920S HAS BEEN CENTRAL TO THE CULTURE OF SOVIET AND LATER RUSSIAN STATE SECURITY ORGANIZATIONS (SEE RID 2020)

60S - 80S 'ACTIVE MEASURES'

“THE HIGH EFFECTIVENESS OF INFORMATION WARFARE SYSTEMS IN COMBINATION WITH HIGHLY ACCURATE WEAPONS AND NONMILITARY MEANS OF INFLUENCE MAKES IT POSSIBLE TO DISORGANIZE THE SYSTEM OF STATE ADMINISTRATION, HIT STRATEGIC INSTALLATIONS AND AFFECT **THE MENTALITY AND MORAL SPIRIT OF THE POPULATION**. IN OTHER WORDS THE EFFECT OF USING THESE MEANS IS COMPARABLE WITH THE DAMAGE RESULTING FROM THE EFFECTS OF WEAPONS OF MASS DESTRUCTION.” CHIEF OF THE RUSSIAN GENERAL STAFF, GENERAL VIKTOR NIKOLAEVICH SAMSONOV (DEC 23 1996)

CONTEMPORARY RUSSIAN MILITARY DOCTRINE EMPHASIZES “APPLYING INFORMATION TECHNOLOGIES FOR THE PRESERVATION OF CULTURAL, HISTORICAL, SPIRITUAL AND MORAL VALUES OF THE MULTI-ETHNIC PEOPLE OF THE RUSSIAN FEDERATION” (2016)

“NEUTRALIZING THE INFORMATION IMPACT INTENDED TO ERODE RUSSIA'S TRADITIONAL MORAL AND SPIRITUAL VALUES.”(2016)

DOCTRINE OF INFORMATION SECURITY OF THE RUSSIAN FEDERATION
[HTTPS://WWW.MID.RU/EN/FOREIGN_POLICY/OFFICIAL_DOCUMENTS/-/ASSET_PUBLISHER/CPTICKB6BZ29/CONTENT/ID/2563163](https://www.mid.ru/en/foreign_policy/official_documents/-/asset_publisher/cptickb6bz29/content/id/2563163)

RUSSIAN/SOVIET INFORMATION WARFARE

THE STORY OF 'THE TRUST' DEPLOYED BY THE CHEKISTS AGAINST WHITE RUSSIAN ÉMIGRÉ OPPOSITION IN THE 1920S HAS BEEN CENTRAL TO THE CULTURE OF SOVIET AND LATER RUSSIAN STATE SECURITY ORGANIZATIONS (SEE RID 2020)

60S - 80S 'ACTIVE MEASURES'

“THE HIGH EFFECTIVENESS OF INFORMATION WARFARE SYSTEMS IN COMBINATION WITH HIGHLY ACCURATE WEAPONS AND **NONMILITARY MEANS OF INFLUENCE MAKES IT POSSIBLE TO DISORGANIZE THE SYSTEM OF STATE ADMINISTRATION, HIT STRATEGIC INSTALLATIONS AND AFFECT THE MENTALITY AND MORAL SPIRIT OF THE POPULATION.** IN OTHER WORDS THE EFFECT OF USING THESE MEANS IS COMPARABLE WITH THE DAMAGE RESULTING FROM THE EFFECTS OF WEAPONS OF MASS DESTRUCTION.” CHIEF OF THE RUSSIAN GENERAL STAFF, GENERAL VIKTOR NIKOLAEVICH SAMSONOV (DEC 23 1996)

CONTEMPORARY RUSSIAN MILITARY DOCTRINE EMPHASIZES “APPLYING INFORMATION TECHNOLOGIES FOR THE PRESERVATION OF CULTURAL, HISTORICAL, SPIRITUAL AND MORAL VALUES OF THE MULTI-ETHNIC PEOPLE OF THE RUSSIAN FEDERATION” (2016)

“NEUTRALIZING THE INFORMATION IMPACT INTENDED TO ERODE RUSSIA'S TRADITIONAL MORAL AND SPIRITUAL VALUES.”(2016)

DOCTRINE OF INFORMATION SECURITY OF THE RUSSIAN FEDERATION
[HTTPS://WWW.MID.RU/EN/FOREIGN_POLICY/OFFICIAL_DOCUMENTS/-/ASSET_PUBLISHER/CPTICKB6BZ29/CONTENT/ID/2563163](https://www.mid.ru/en/foreign_policy/official_documents/-/asset_publisher/cptickb6bz29/content/id/2563163)

RUSSIAN/SOVIET INFORMATION WARFARE

THE STORY OF 'THE TRUST' DEPLOYED BY THE CHEKISTS AGAINST WHITE RUSSIAN ÉMIGRÉ OPPOSITION IN THE 1920S HAS BEEN CENTRAL TO THE CULTURE OF SOVIET AND LATER RUSSIAN STATE SECURITY ORGANIZATIONS (SEE RID 2020)

60S - 80S 'ACTIVE MEASURES'

“THE HIGH EFFECTIVENESS OF INFORMATION WARFARE SYSTEMS IN COMBINATION WITH HIGHLY ACCURATE WEAPONS AND **NONMILITARY MEANS OF INFLUENCE MAKES IT POSSIBLE TO DISORGANIZE THE SYSTEM OF STATE ADMINISTRATION, HIT STRATEGIC INSTALLATIONS AND AFFECT THE MENTALITY AND MORAL SPIRIT OF THE POPULATION.** IN OTHER WORDS THE EFFECT OF USING THESE MEANS IS COMPARABLE WITH THE DAMAGE RESULTING FROM THE EFFECTS OF WEAPONS OF MASS DESTRUCTION.” CHIEF OF THE RUSSIAN GENERAL STAFF, GENERAL VIKTOR NIKOLAEVICH SAMSONOV (DEC 23 1996)

CONTEMPORARY RUSSIAN MILITARY DOCTRINE EMPHASIZES “APPLYING INFORMATION TECHNOLOGIES FOR **THE PRESERVATION OF CULTURAL, HISTORICAL, SPIRITUAL AND MORAL VALUES** OF THE MULTI-ETHNIC PEOPLE OF THE RUSSIAN FEDERATION” (2016)

“NEUTRALIZING THE INFORMATION IMPACT INTENDED TO ERODE **RUSSIA'S TRADITIONAL MORAL AND SPIRITUAL VALUES.**”(2016)

DOCTRINE OF INFORMATION SECURITY OF THE RUSSIAN FEDERATION
[HTTPS://WWW.MID.RU/EN/FOREIGN_POLICY/OFFICIAL_DOCUMENTS/-/ASSET_PUBLISHER/CPTICKB6BZ29/CONTENT/ID/2563163](https://www.mid.ru/en/foreign_policy/official_documents/-/asset_publisher/cptickb6bz29/content/id/2563163)

READING RUSSIAN MILITARY DOCTRINE

- This emphasis on moral and spiritual values in the *Doctrine of Information Security of the Russian Federation* has puzzled commentators including Daniel Bagge (2019) Timothy Thomas (2016) and others
- It would be a mistake to think of this solely as a pretext to permit overly broad censorship within Russia.
 - Russian intellectuals and military figures are correctly convinced of the central role of culture (moral and spiritual values) in interstate conflict.

THEORETICAL STARTING POINT AND RECOMMENDATIONS

OVERVIEW

- Social infrastructure is part of contemporary warfare
- Social norms are constitutive of institutions
- Institutions can be undermined by undermining social norms
- We know how (in broad strokes) to intervene in and defend social norms

RECOMMENDATIONS

- rethink the social attack surface in social terms rather than in terms of individual psychology
- increased research into social vulnerabilities
- consider the role of choice architectures in the cyber-social interface

THANK YOU

STACY ELMER
SLELMER@KU.EDU

JOHN SYMONS
JOHNSYMONS@KU.EDU



1912



1920



1923



1929



1941



1946



2006-now