Towards a model of trust drawn from neuroscience, psychology, and economics:

Implications for the New and Next Generation Computing Environments

Frank Krueger, Ph.D.

Associate Professor of
Systems Social Neuroscience

School of Systems Biology
George Mason University, VA, USA

Chief, Social Cognition and Interaction:
Functional Imaging (SCI:FI) Lab

Director, Transdisciplinary Research Union
for the Study of Trust (T-R-U-S-T)

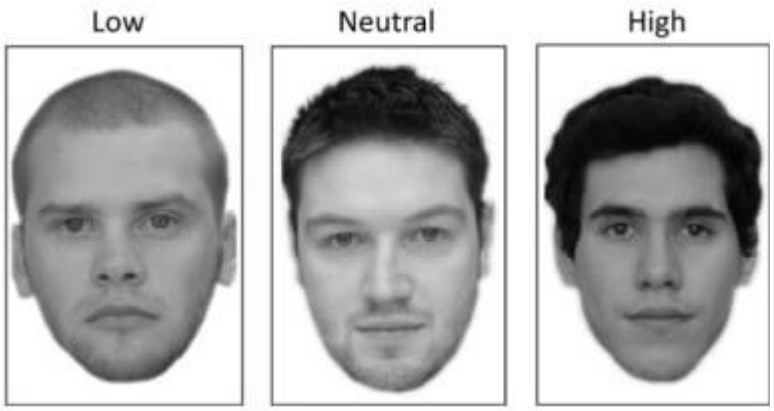
SCI:FI LABORATORY

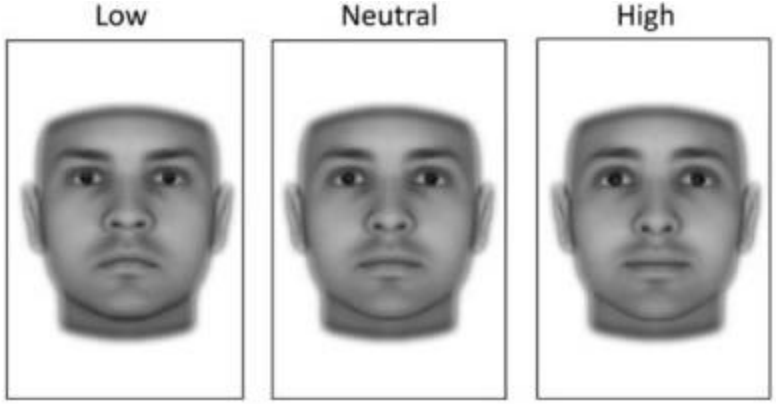
CROSSING Conference 2019
Darmstadt 9|10|19

To Trust or Not to Trust: That is the Question

How much money would you lend this person?

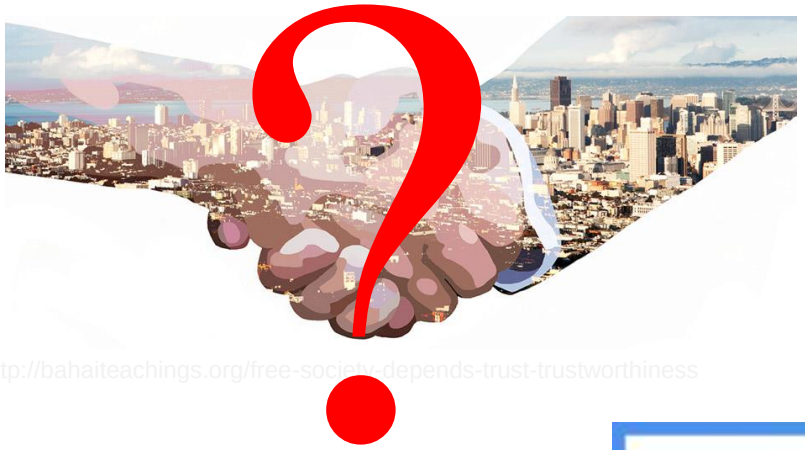
$0 $2 $6 $8 $10



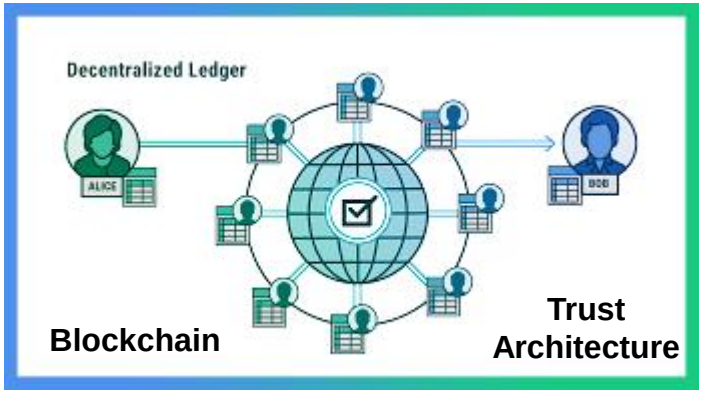


Why Should We Care About Trust ?

http://www.handicappeddoctor.com/?p=744

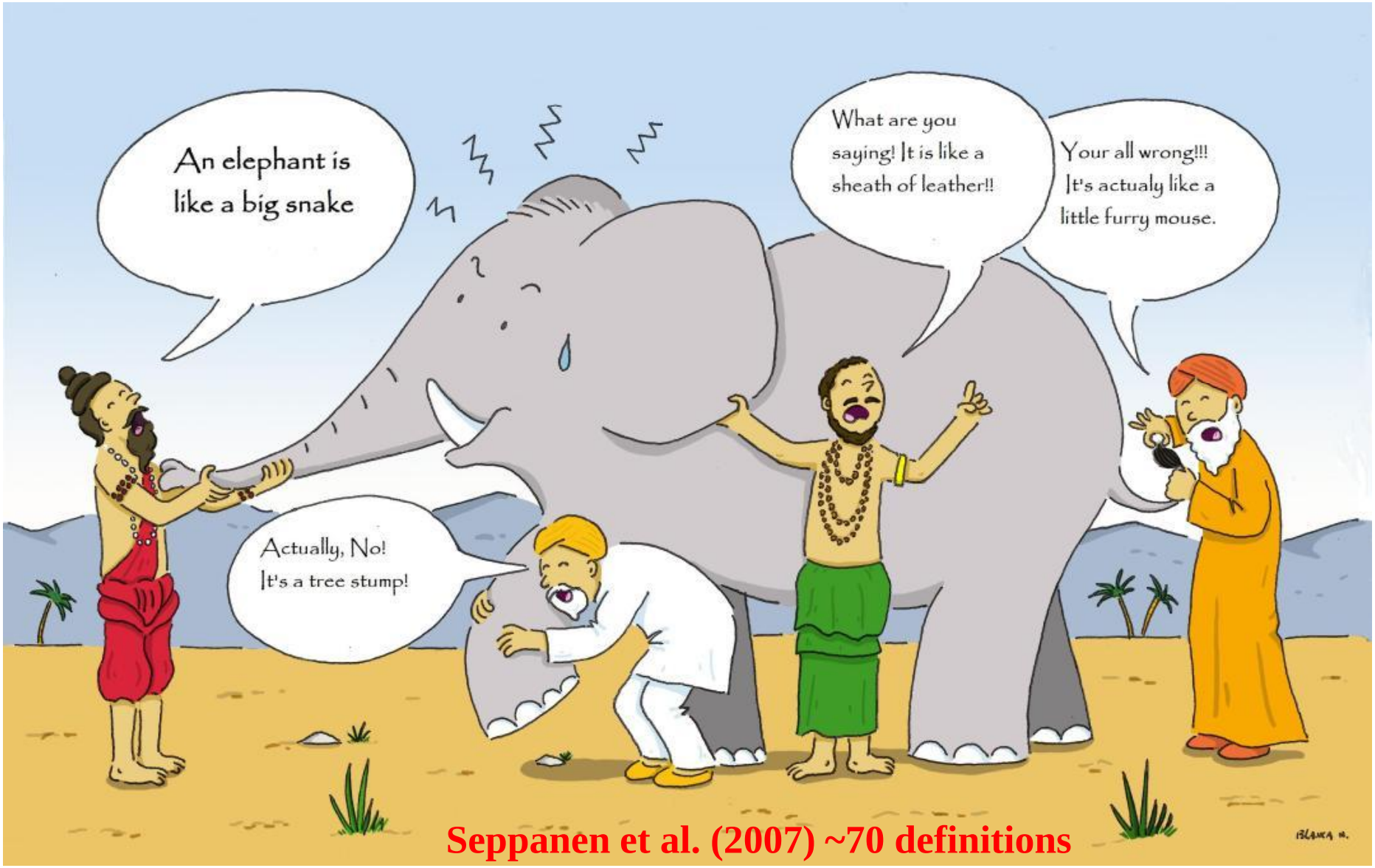


http://bahaiteachings.org/free-society-depends-trust-trustworthiness



https://medium.com/nhct-nanohealth-care-token/decentralized-vs-centralized-systems-cb31b95928c5

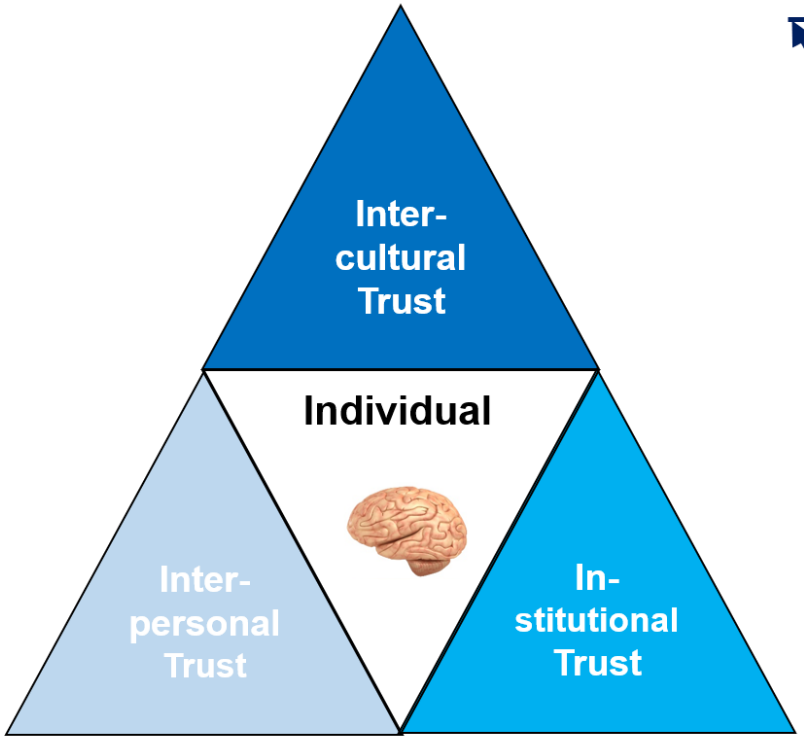
Parable: The Blind Men and the Elephant

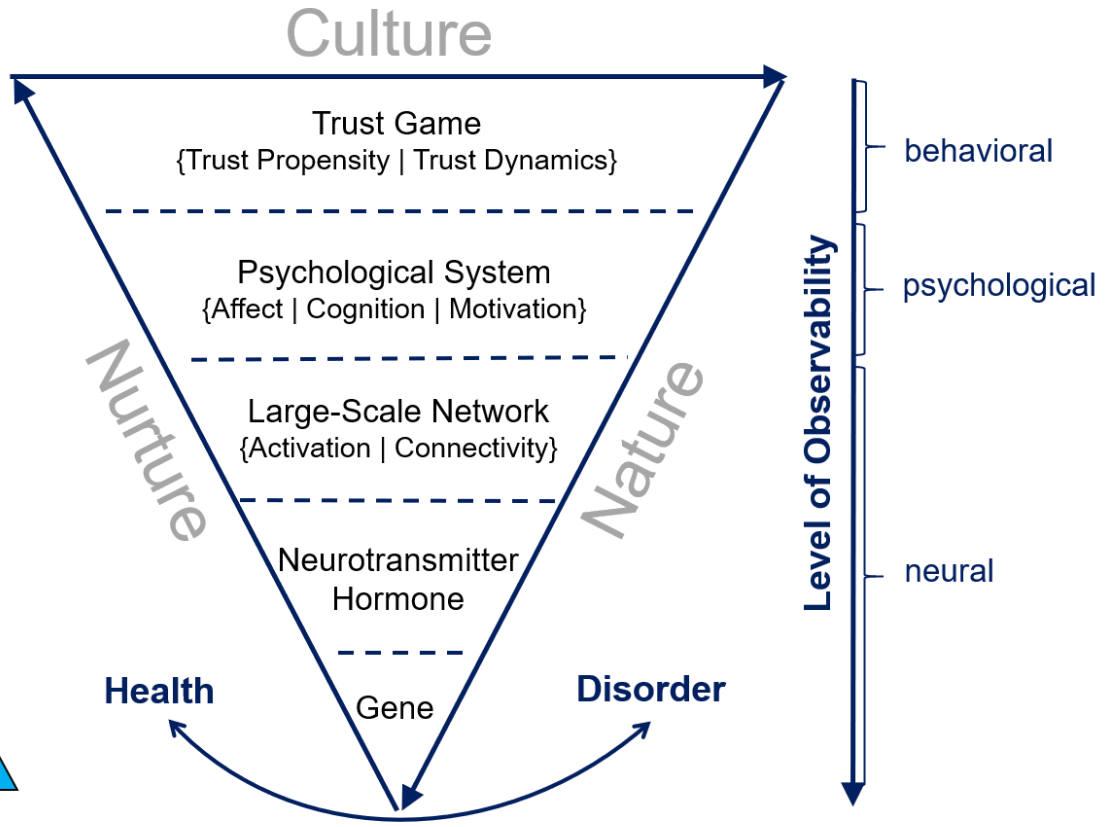


Seppanen et al. (2007) ~70 definitions

https://equilibregaia.files.wordpress.com/2011/02/conte-sufc3ad0001-copiaeng.jpg?w=1075

NeuroPsychoEconomic Approach





Definition → Framework → Model

Definition & Paradigm

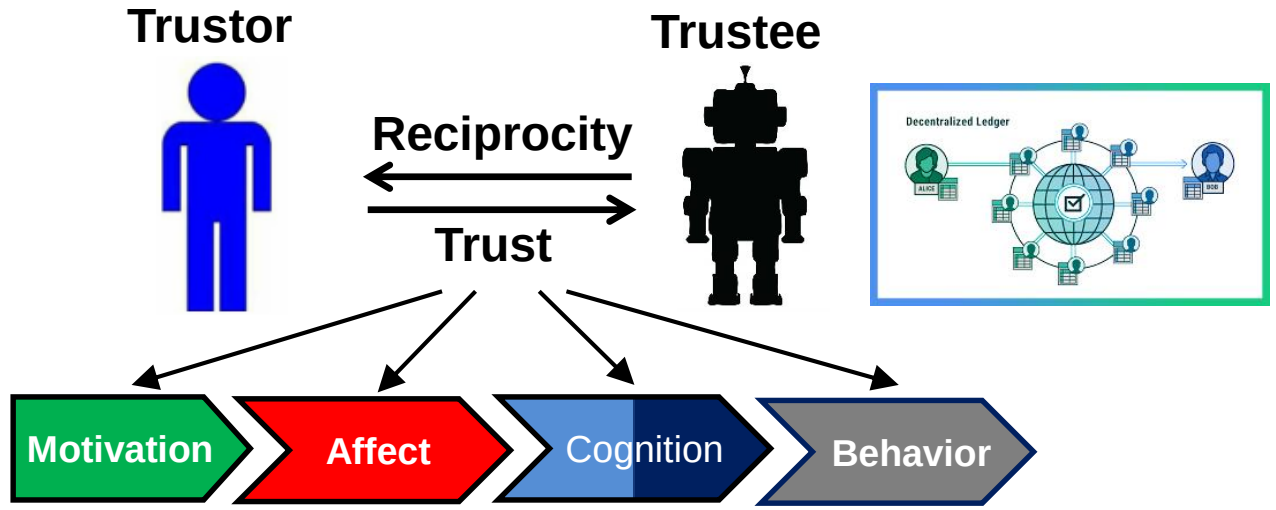
https://www.theodysseyonline.com/trust-underestimated-fear

Interpersonal trust encompasses one‘s willingness to accept **vulnerability** based on the expectation regarding the behavior of another party that will produce some positive outcome in the future.

Working Definition of Trust

Trust in <u>reciprocity</u> entails a psychological state in which a **trustor** is willing to be vulnerable to the risk of **treachery** (**affect**) based on the **expectation** (**cognition**) regarding the action of a **trustee** that will produce some anticipated **reward** (**motivation**) due to reciprocation in the future.

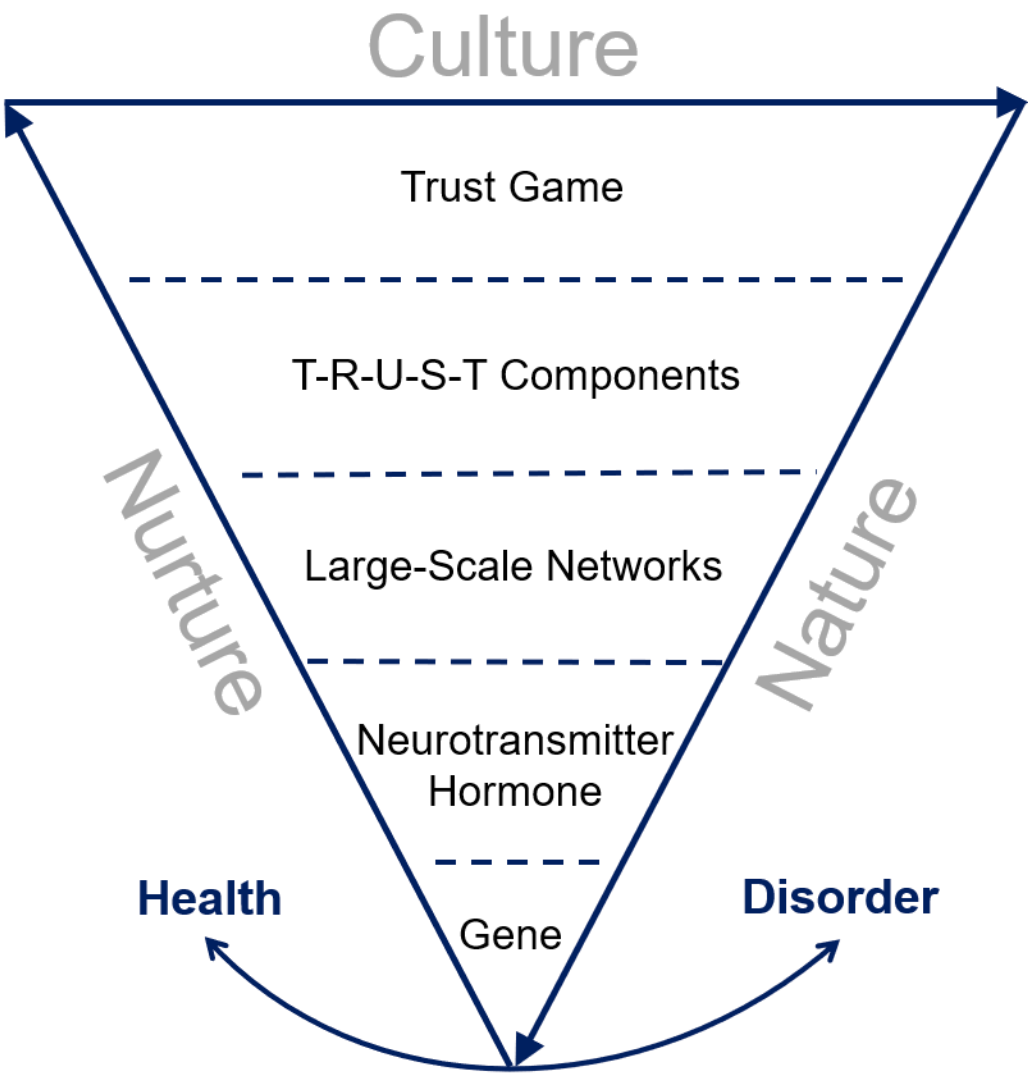
Social Dilemma
(uncertainty)

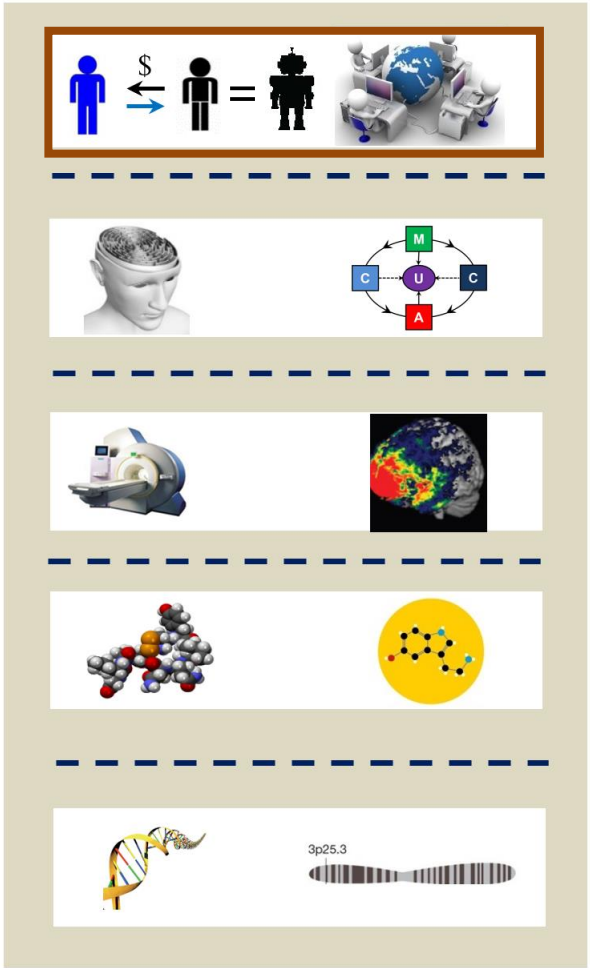


T-R-U-S-T

Treachery **Trustworthiness**

NeuroPsychoEconomic Framework





Behavior: Trust Game

Game Theory:

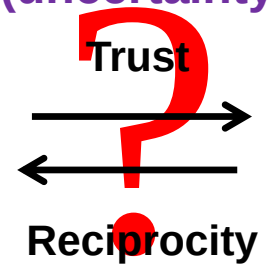
Risk

Trustor

A

Social Dilemma
(uncertainty)

Trust



Reciprocity

3*A
B(A)

Trustee

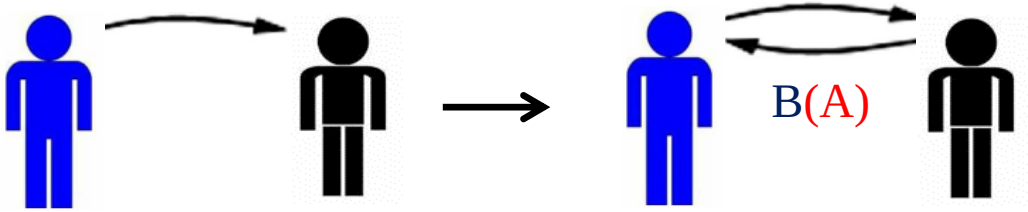
vs.

Computer

Bounded Rationality

Berg et al., 1995

Stages:



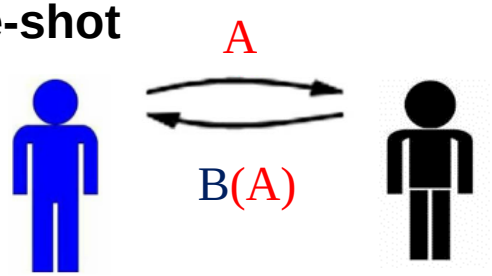
(i) Trust

(iii) Feedback

(ii) Reciprocity

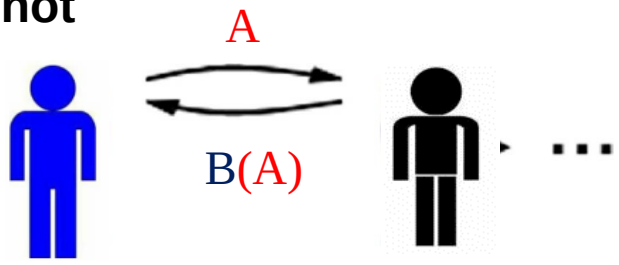
Types:

one-shot



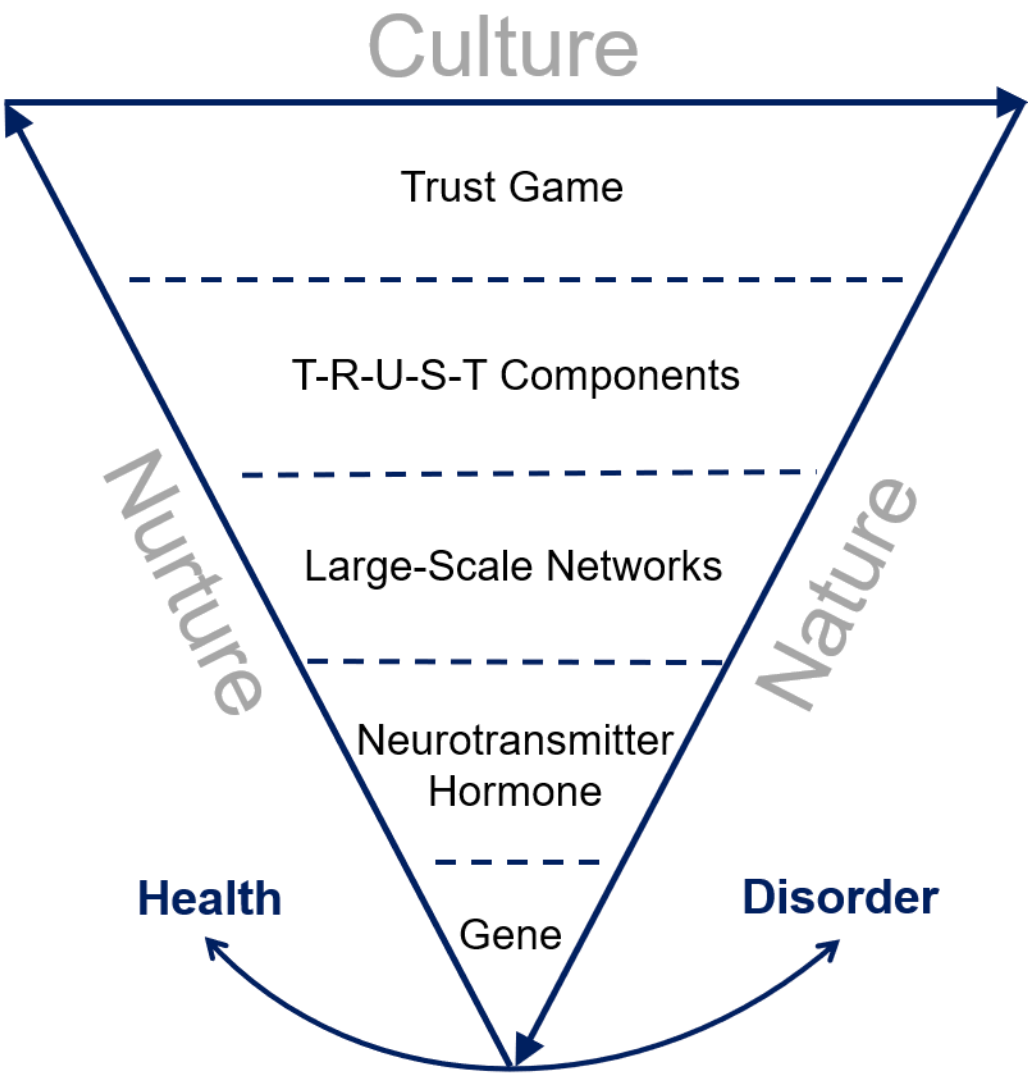
Trust Propensity

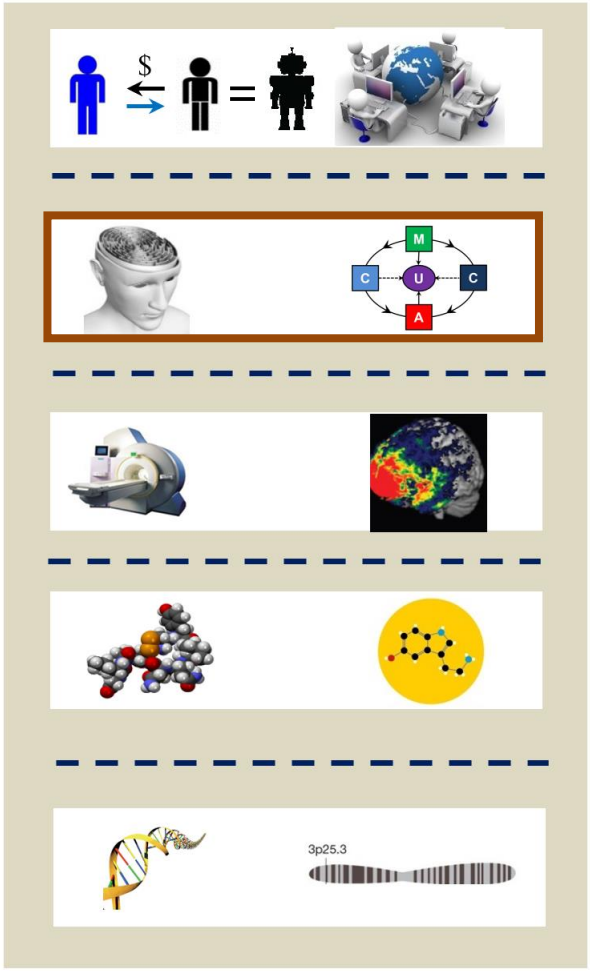
multi-shot



Trust Dynamics

NeuroPsychoEconomic Framework





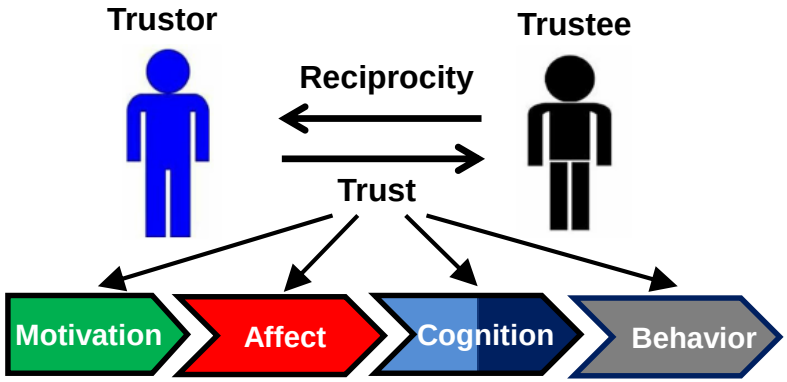
Psyche: T-R-U-S-T Components

T-R-U-S-T Model

Formation **Resolution**

T – R –> U <– S – T

Social Dilemma (uncertainty)

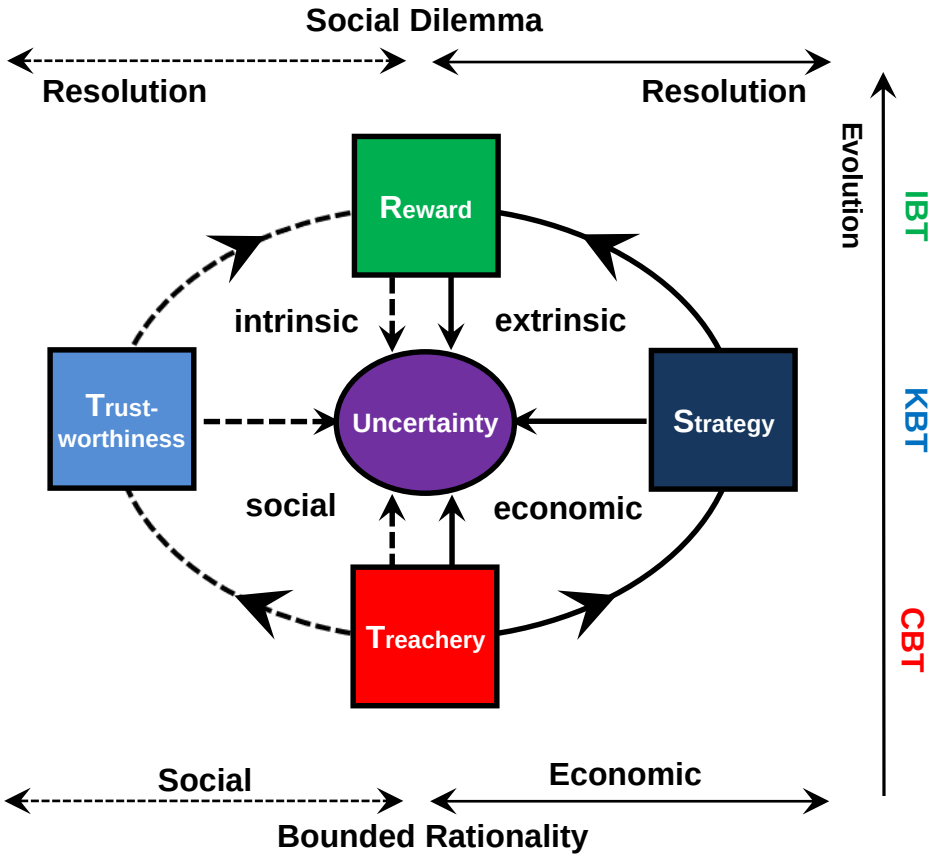


Bounded Rationality

CBT, calculus-based trust

KBT, knowledge-based trust

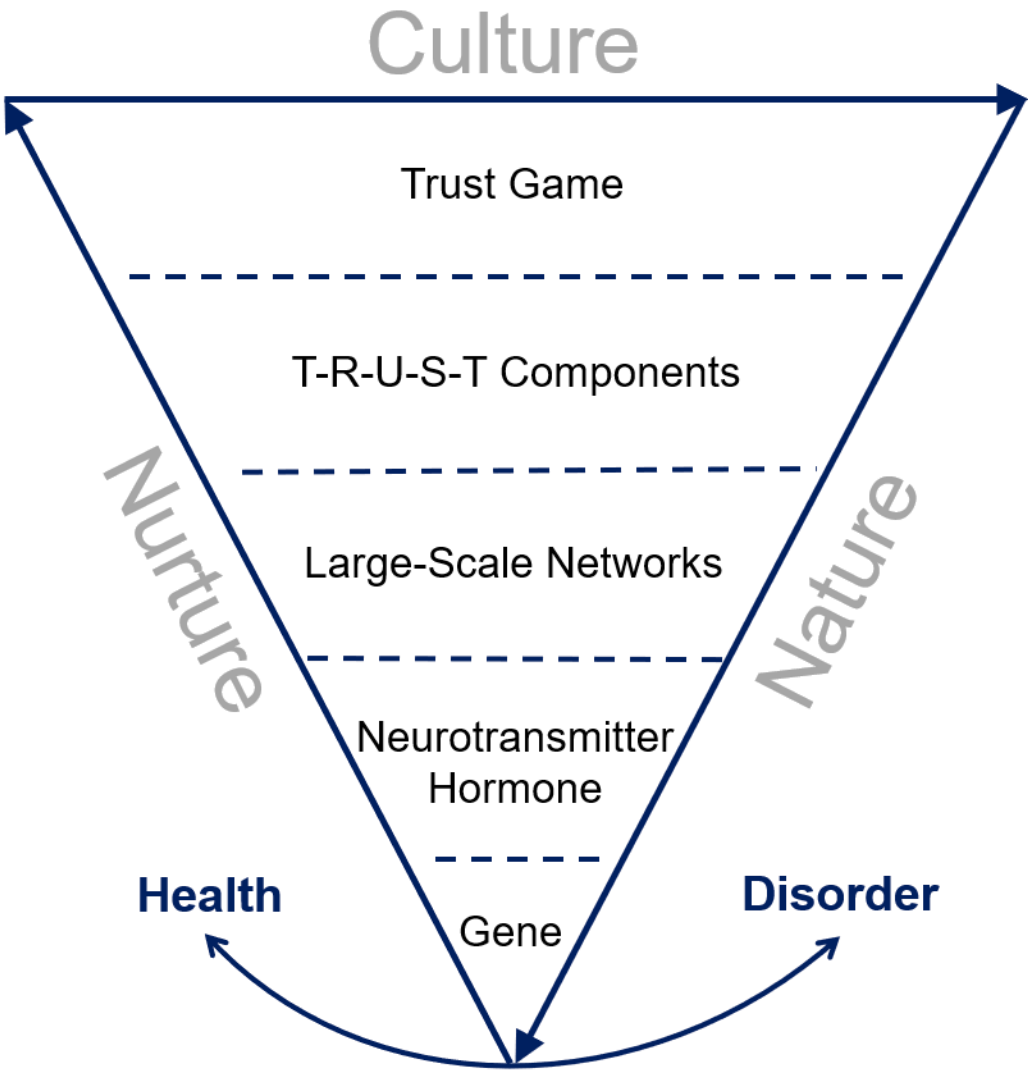
IBT, identification-based trust

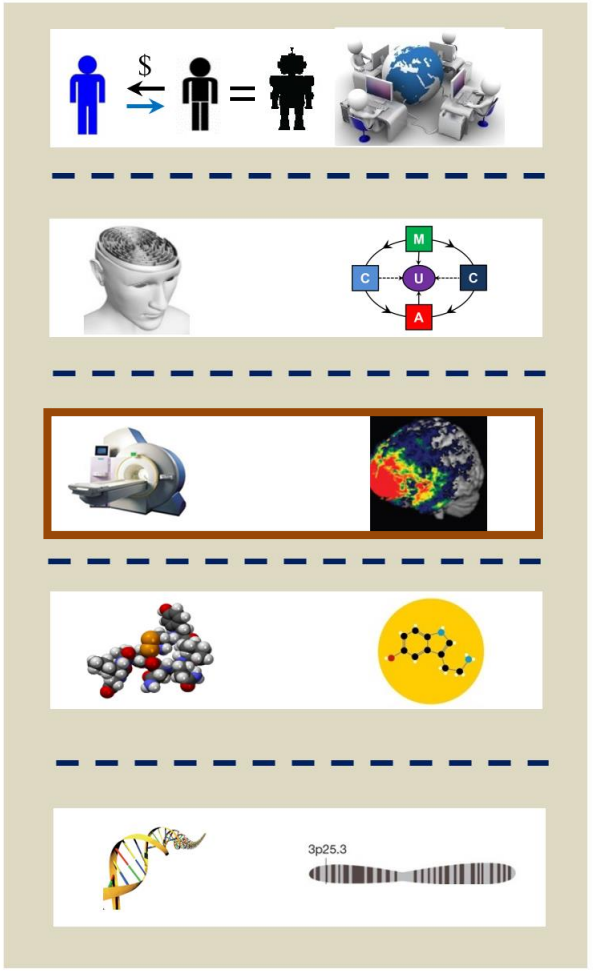


Lewicki & Bunker (1995)

Krueger & Meyer-Lindenberg (TINS, 2019)

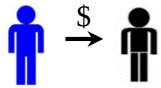
NeuroPsychoEconomic Framework





Functional Neuroimaging

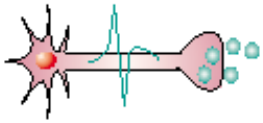
Blood Oxygen Level Dependent (BOLD) Signal



(1) Neuronal activity

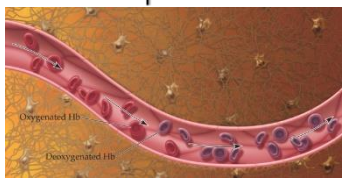
Stimulus

or modulation in background activity

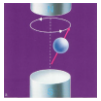


(2) Neurovascular coupling

(3) Haemodynamic response



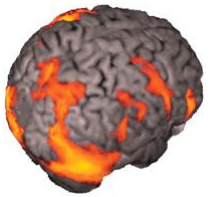
glucose
oxygen



T1,T2

(4) Detection by MRI scanner

fMRI BOLD response

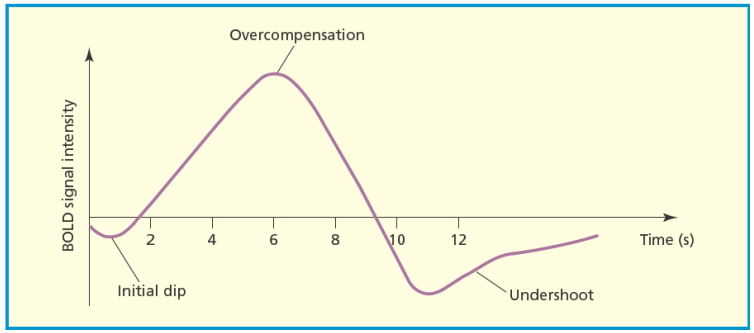


- Excitatory activity and inhibitory activity

- Metabolic signal unknown

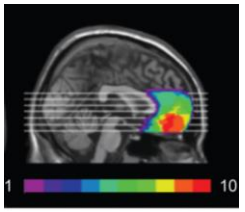
- Blood flow
- Blood oxygenation level

- Magnetic field strength
- TR, repetition time
- TE, echo time
- Spin or gradient echo EPI



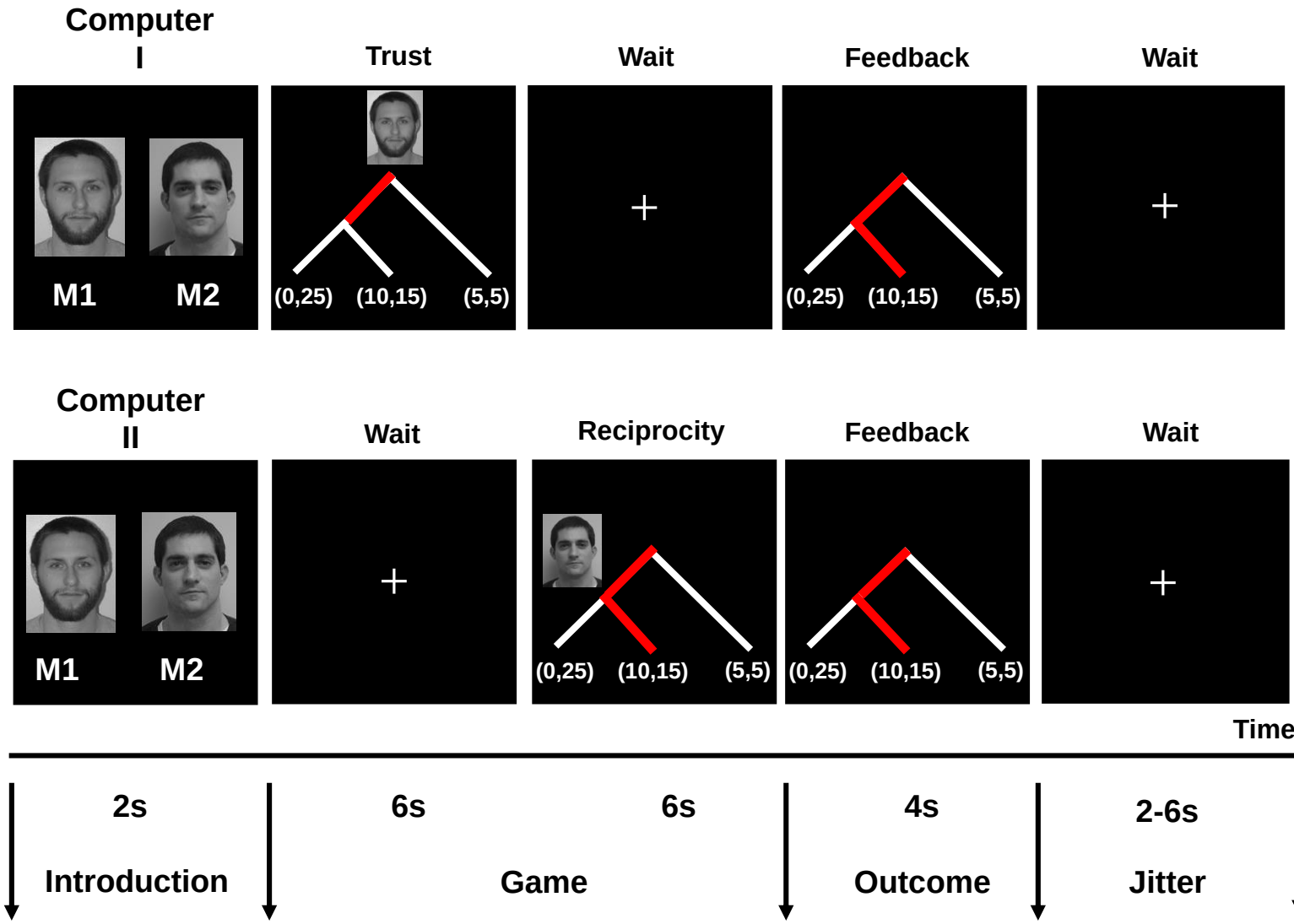
1) Control Condition

2) Lesion Evidence



↑ neural activity ➔ ↑ blood flow ➔ ↑ de-/ oxygenated Hgb ➔ ↑ T2* ➔ ↑ MR signal

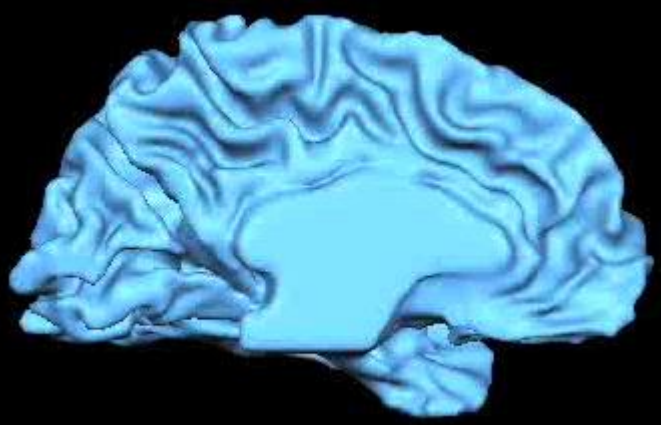
Task-Based Functional MRI



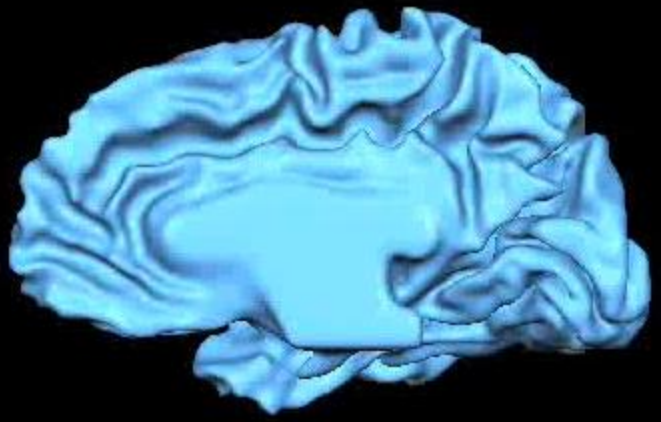
Krueger et al., 2007

Inside the Scanners

Scanner I

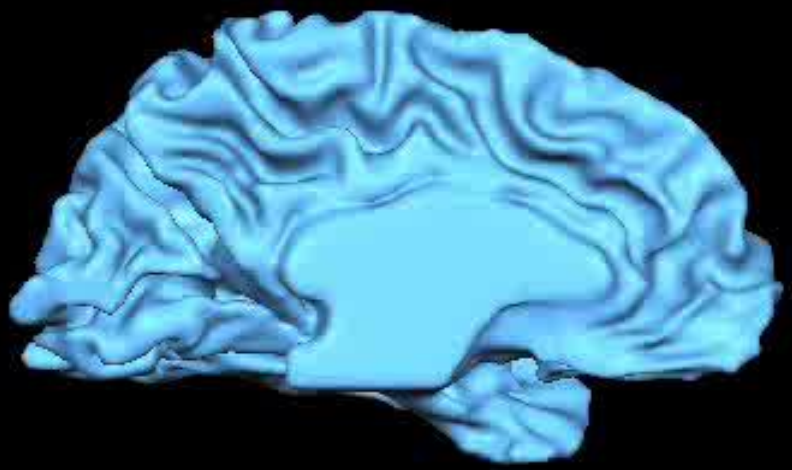


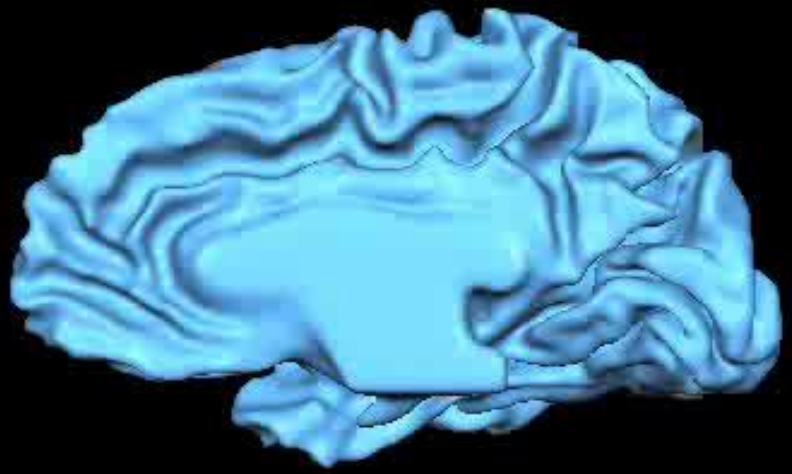
Scanner II



M1 M2

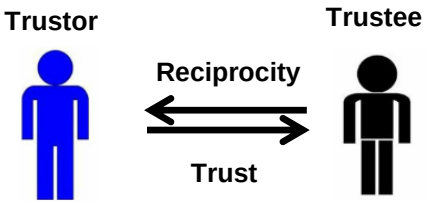
M1 M2



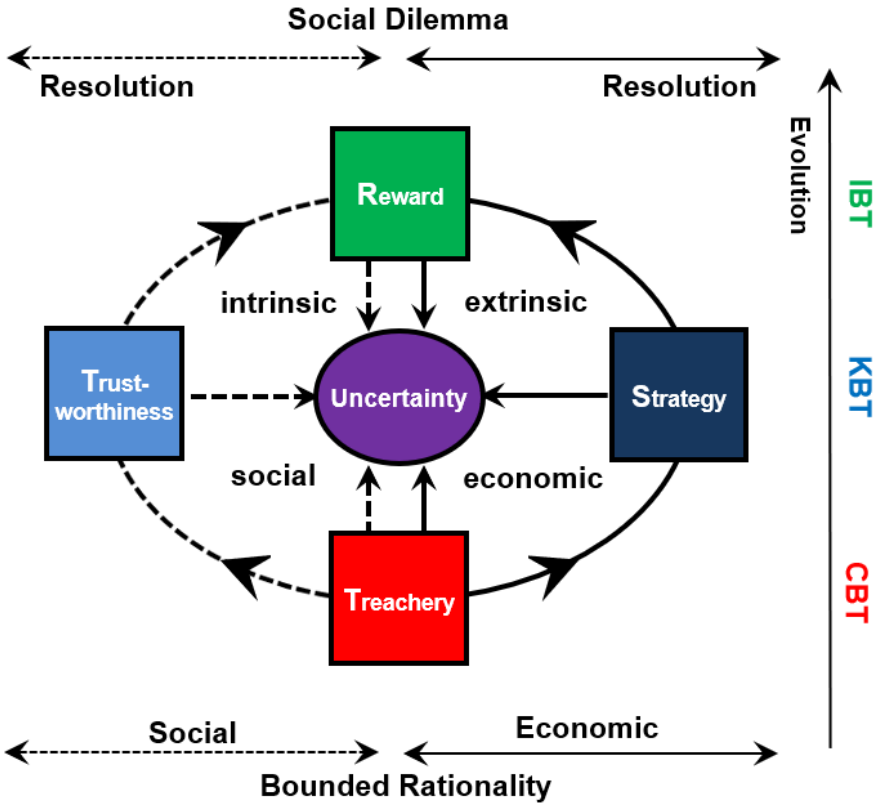


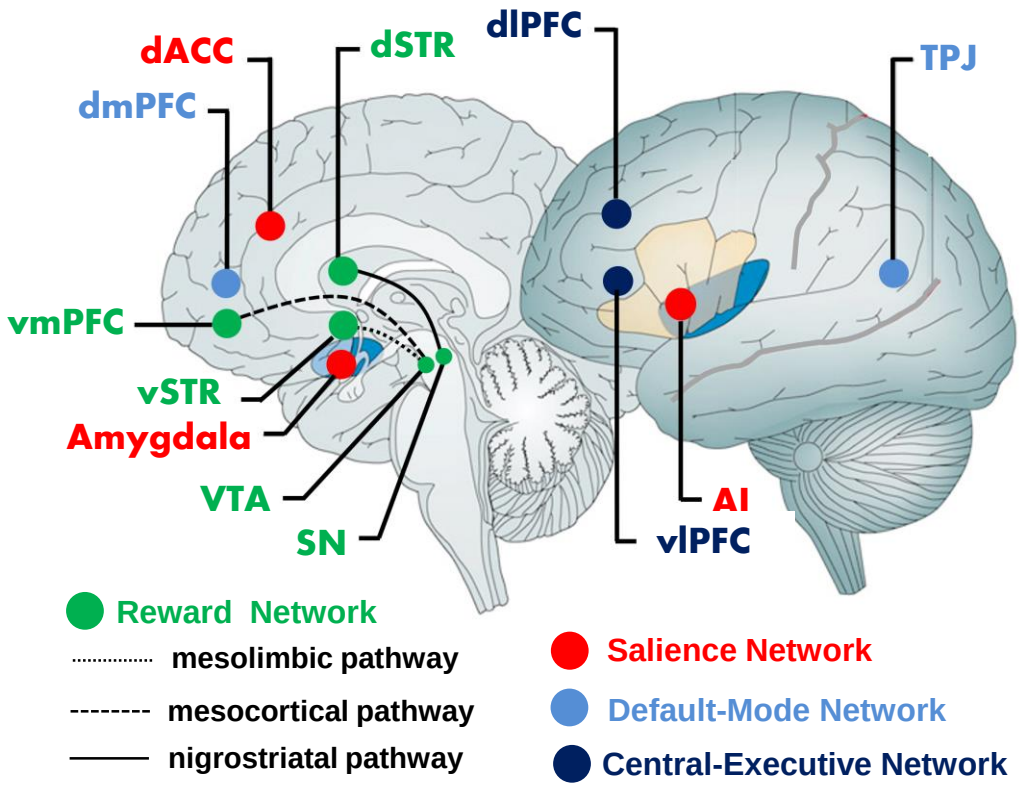
Brain: Large-Scale Networks

T – R – U – S – T



Domain-General Large-Sale Networks



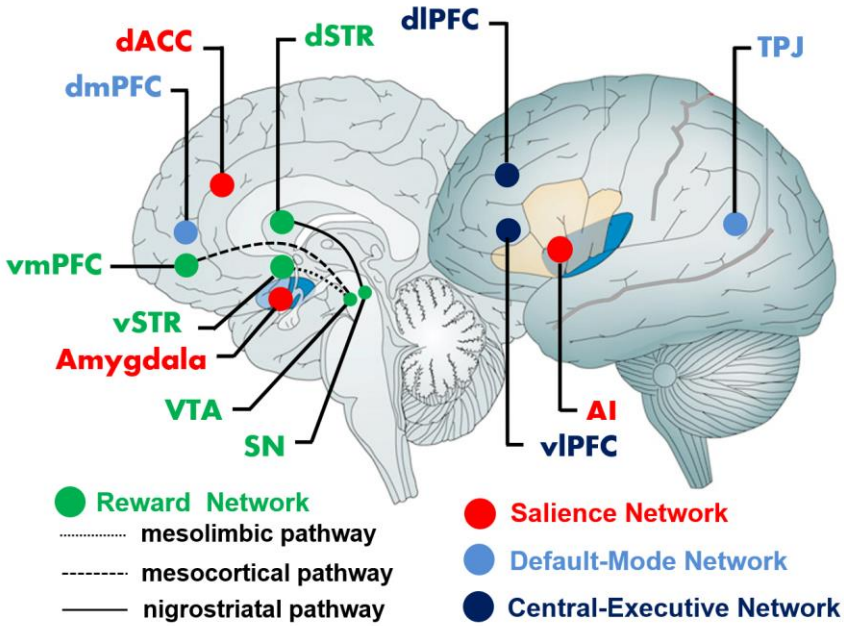


CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust; **VTA**, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoparietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC

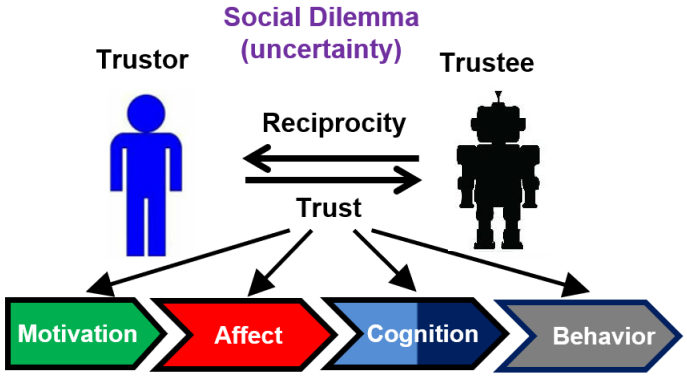
Social Dilemma: Formation

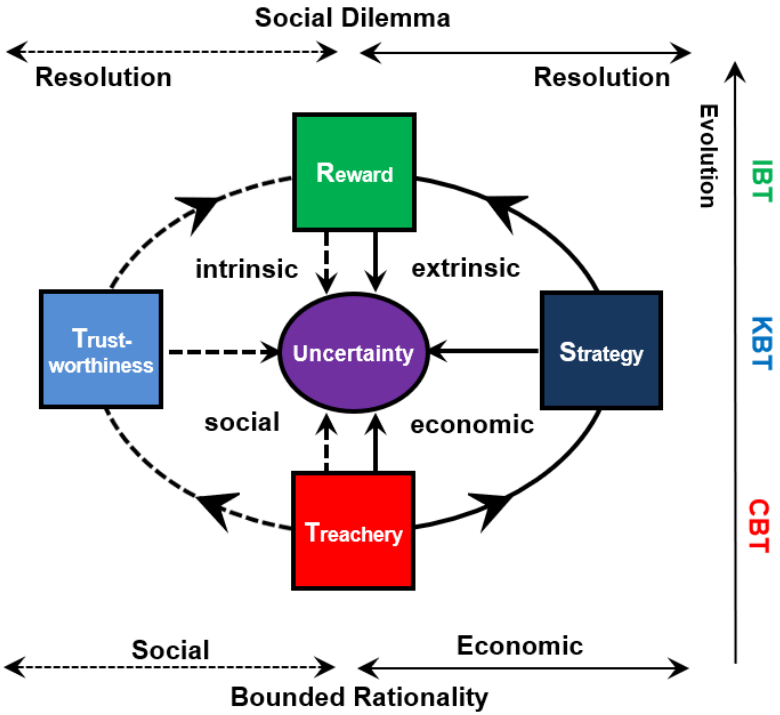
T – R – U – S – T

Formation **Resolution**



VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC



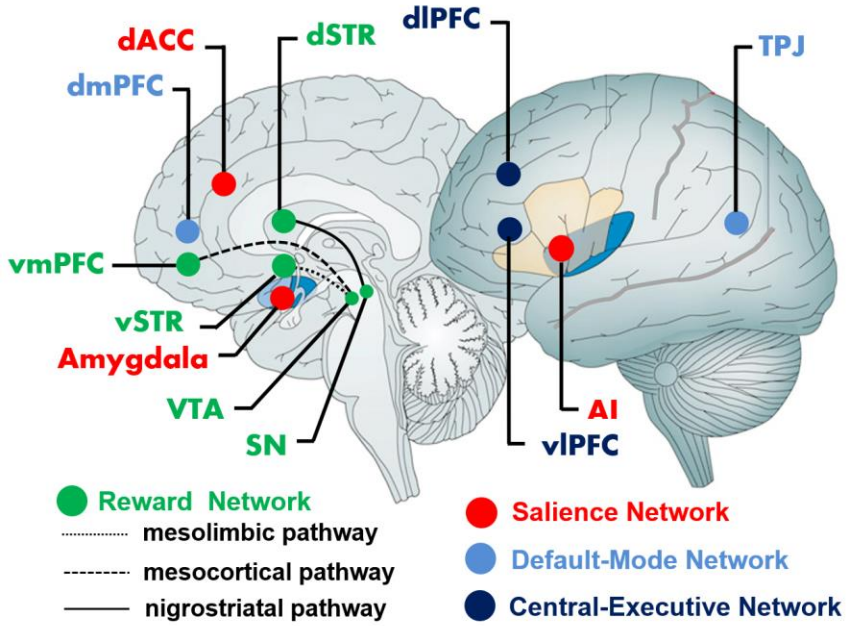


CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust;

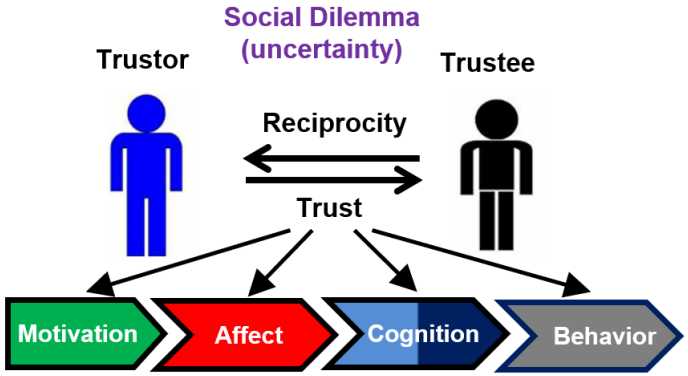
Reward Network = Reward

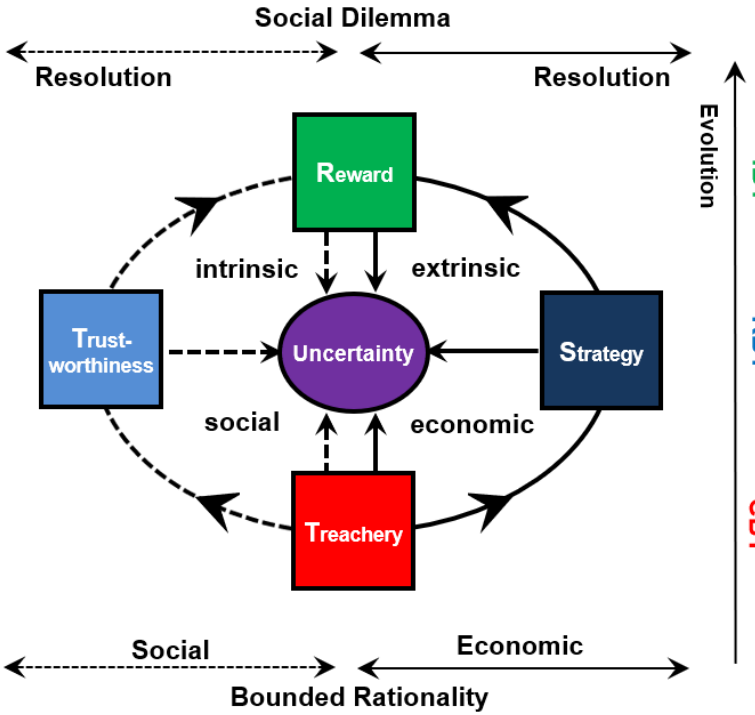
Anticipation of Reward

- **Subjective Reward Value (Utility)**
  - Ventromedial prefrontal cortex (vmPFC)
- **Reinforcement Learning**
  - Striatum (STR)



VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC



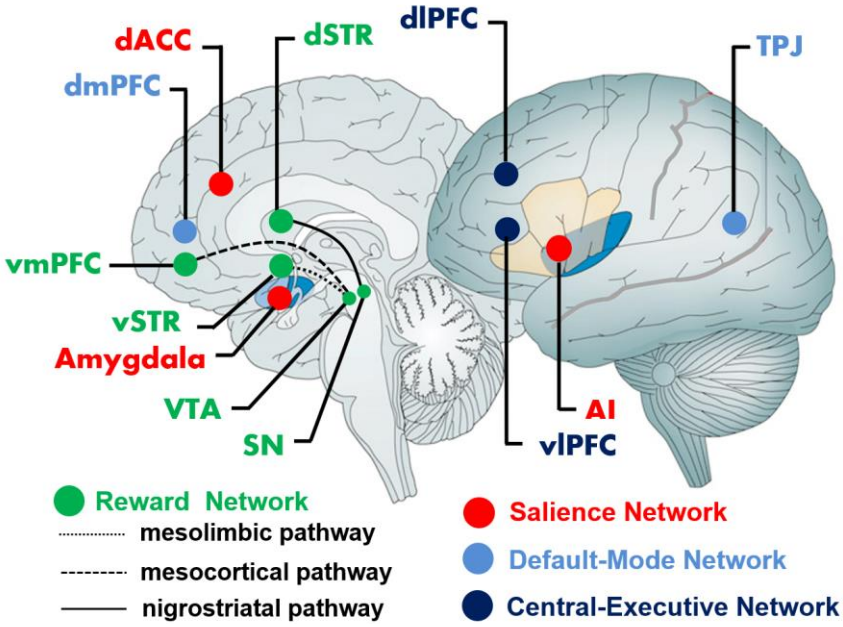


CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust;

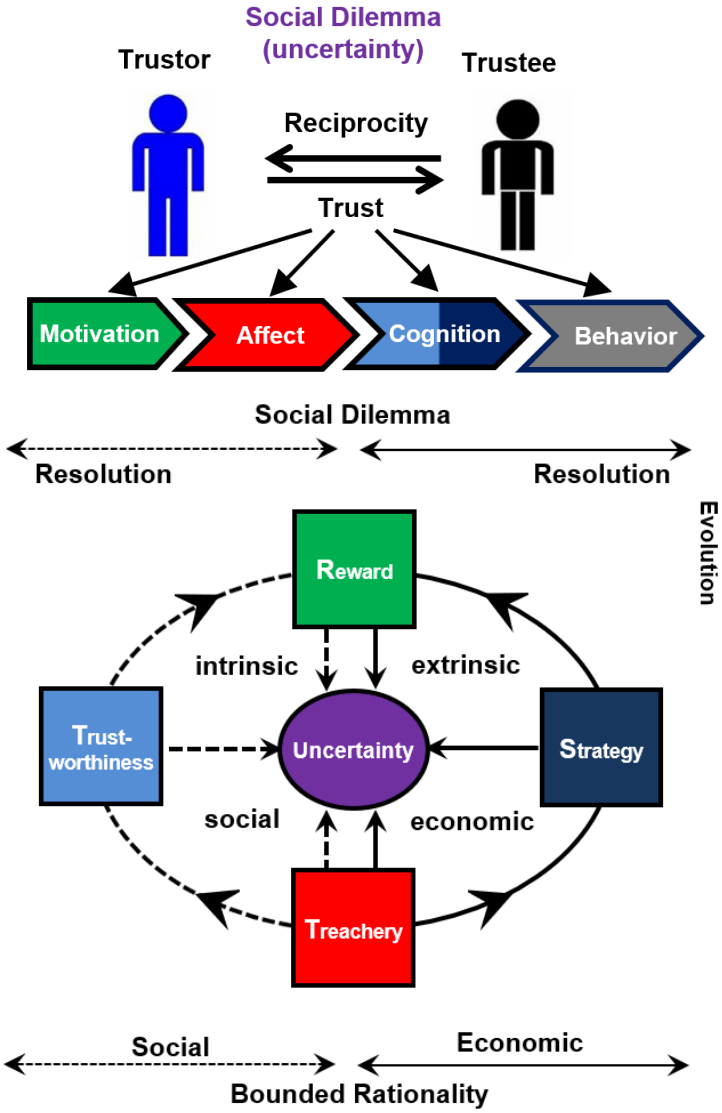
Salience Network = Treachery

Risk of Treachery

- **Threat of Treachery (Emotional Response)**
  - Amygdala
- **Treachery Aversion (Subjective Feeling)**
  - Anterior Insula (AI)
- **Treachery vs. Reward (Monitoring)**
  - Dorsal Anterior Cingulate Cortex (dACC)



VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC

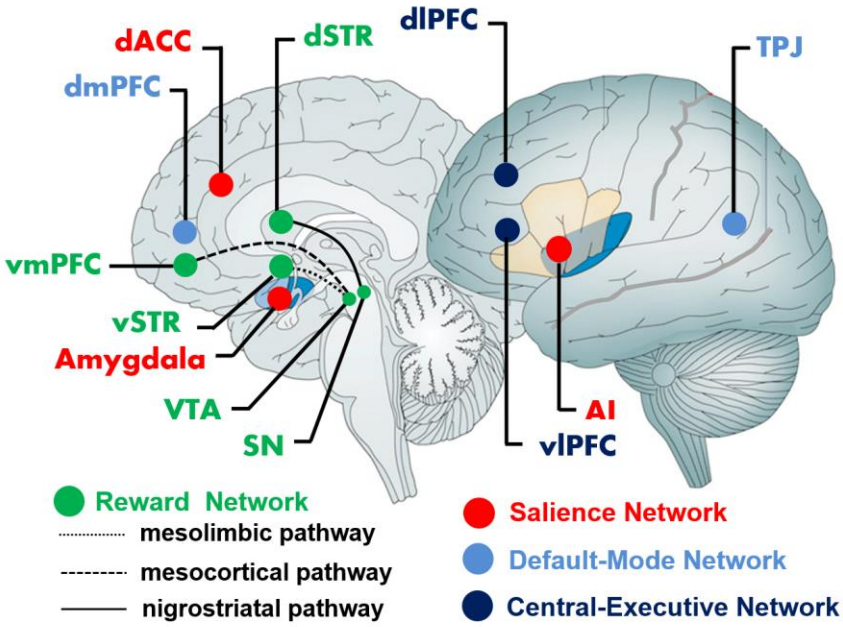


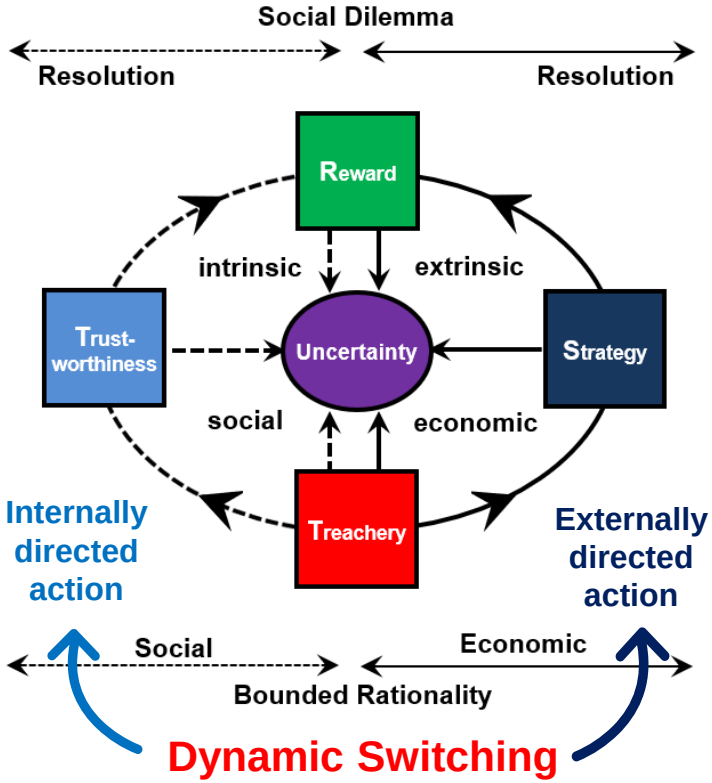
CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust;

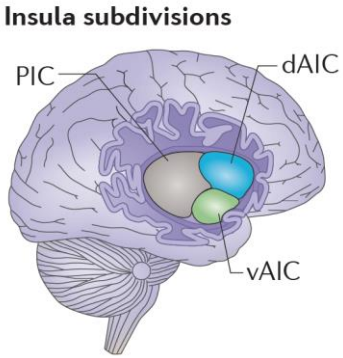
Social Dilemma: Resolution

T – R – U – S – T

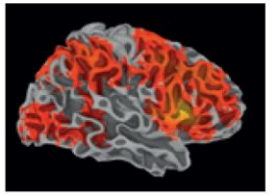
Formation **Resolution**



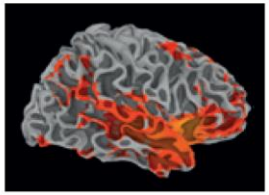




dAIC co-activation associated with cognitive processing areas



vAIC co-activation associated with affective processing areas



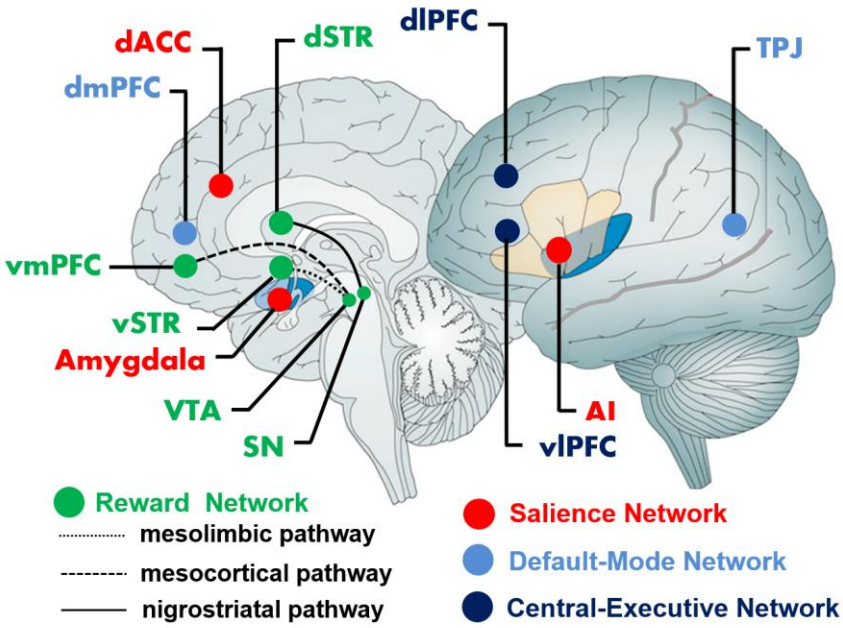
VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC

Uddin (2014)

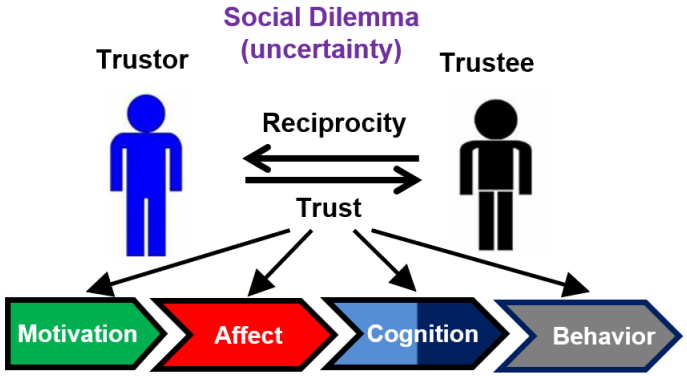
Default-Mode Network = Trustworthiness

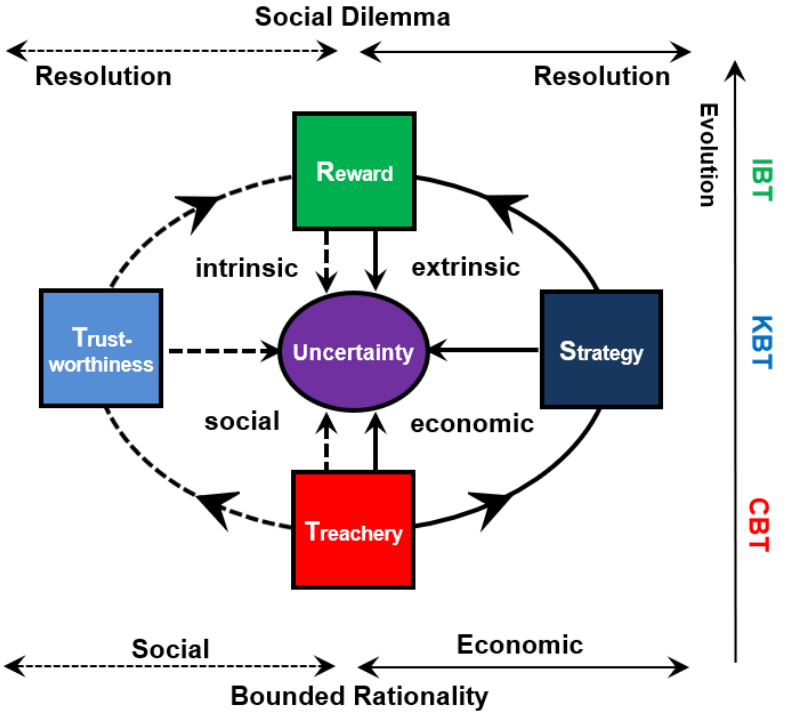
Evaluation of Trustworthiness

- **Inferring and Attributing of Traits**
  - Dorsomedial prefrontal cortex (dmPFC)
- **Inferring and Attributing of Intentions**
  - Temporoparietal junction (TPJ)



VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC



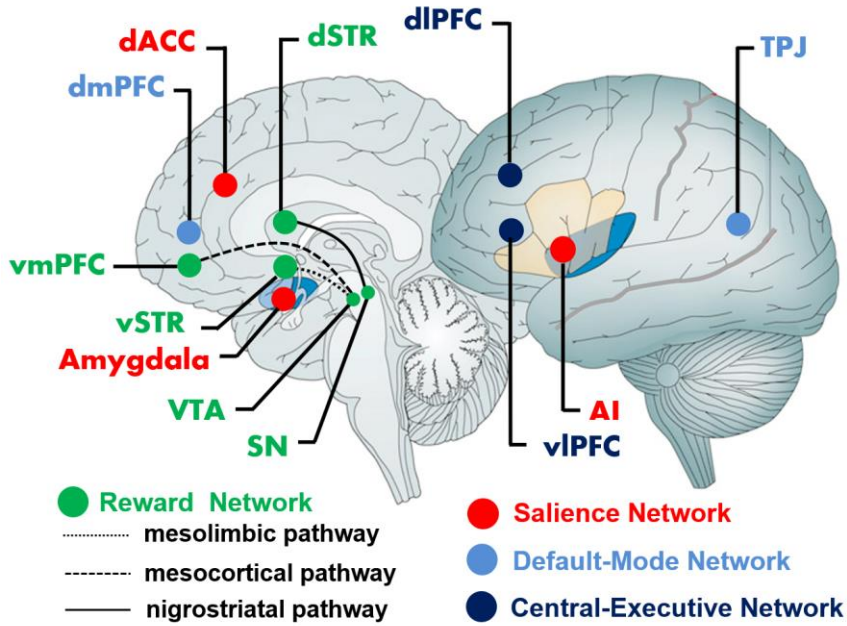


CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust;

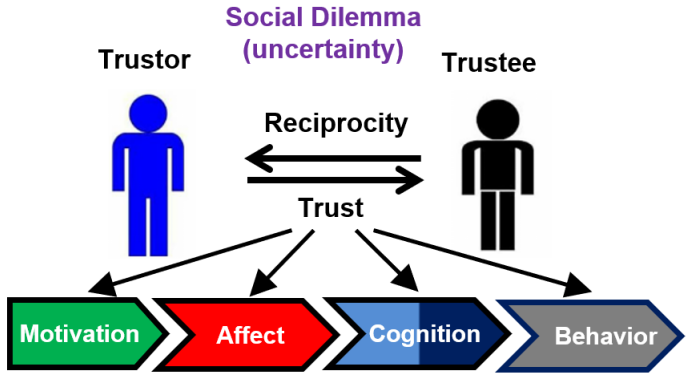
Central-Executive Network = Strategy

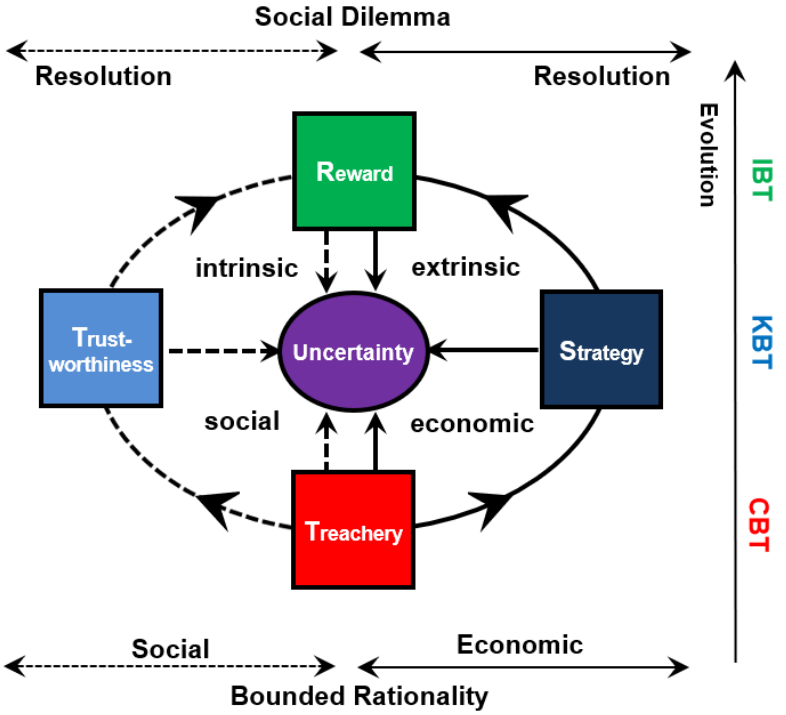
Context-Based Adaption of Strategy

- **Accounting Conflicting Evidence**
  - Dorsolateral PFC (dlPFC)
- **Disaccounting Conflicting Evidence**
  - Ventrolateral PFC (vlPFC)



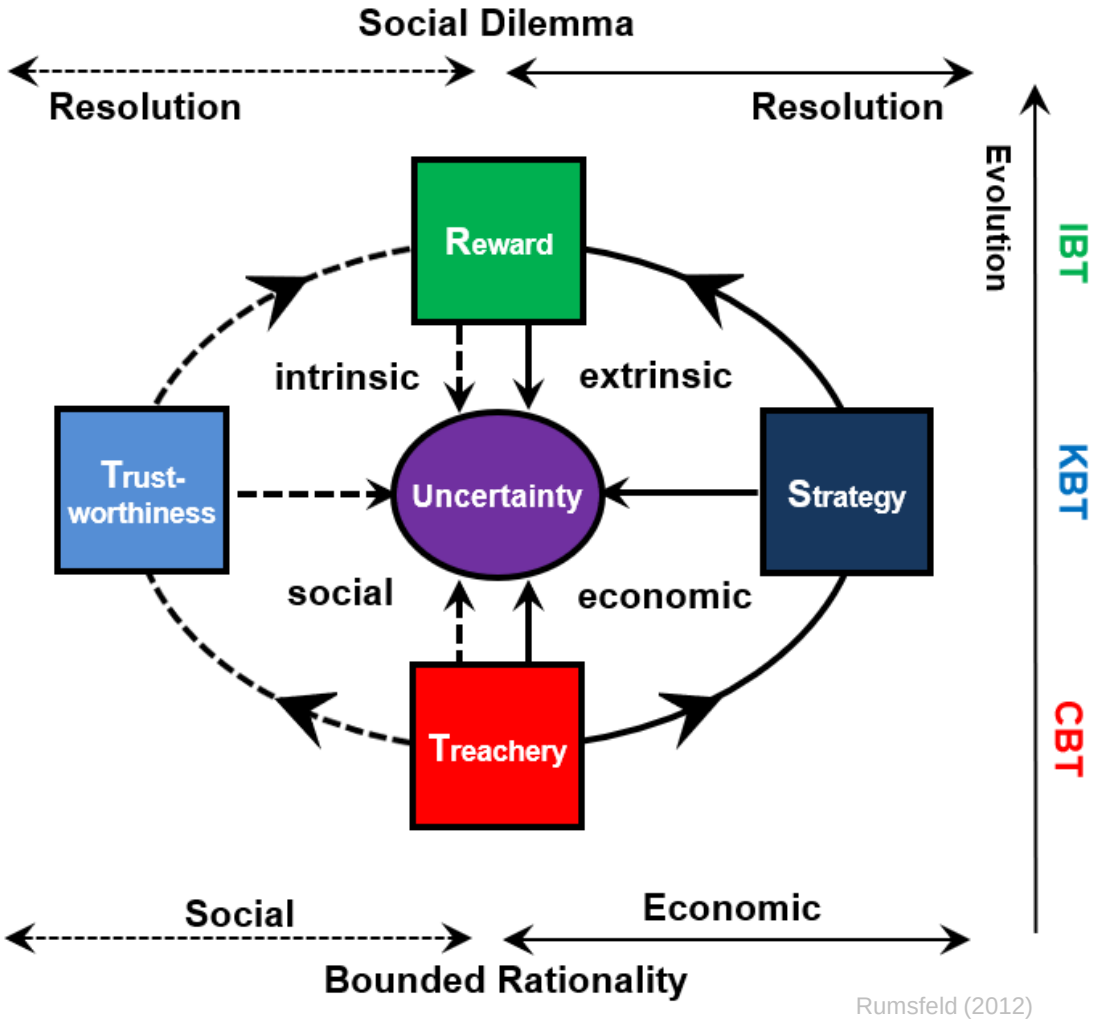
VTA, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoarietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC





CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust;

Types of Trust



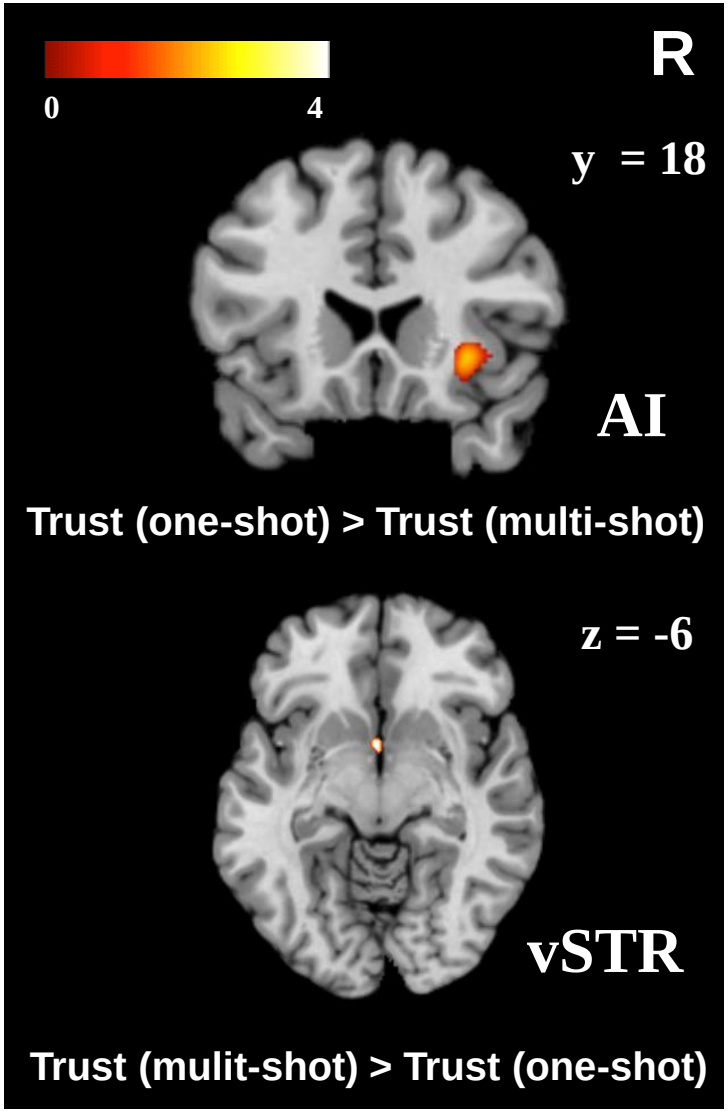
Rumsfeld (2012)

CBT, calculus-based trust [The unknown unknowns]

KBT, knowledge-based trust [The known unknowns]

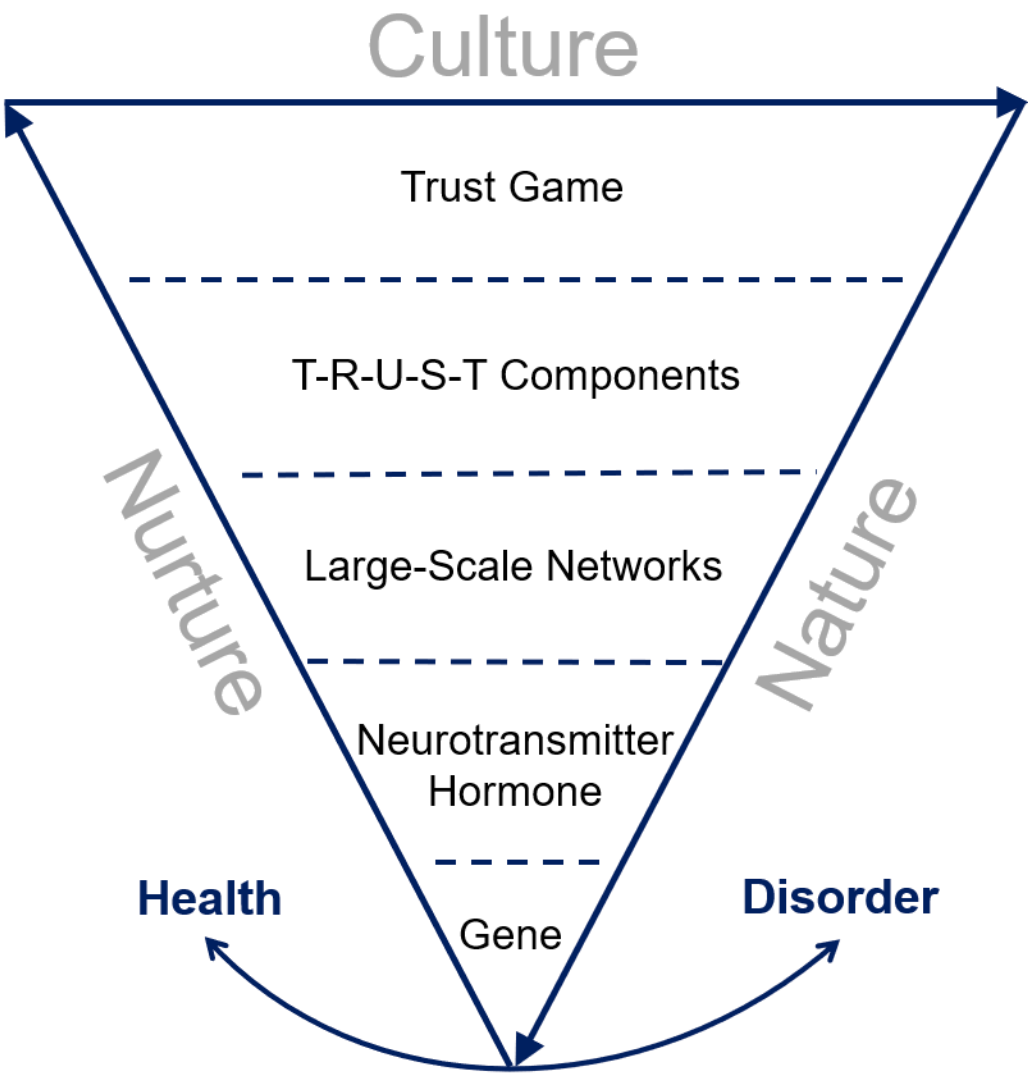
IBT, identification-based trust [The known knowns]

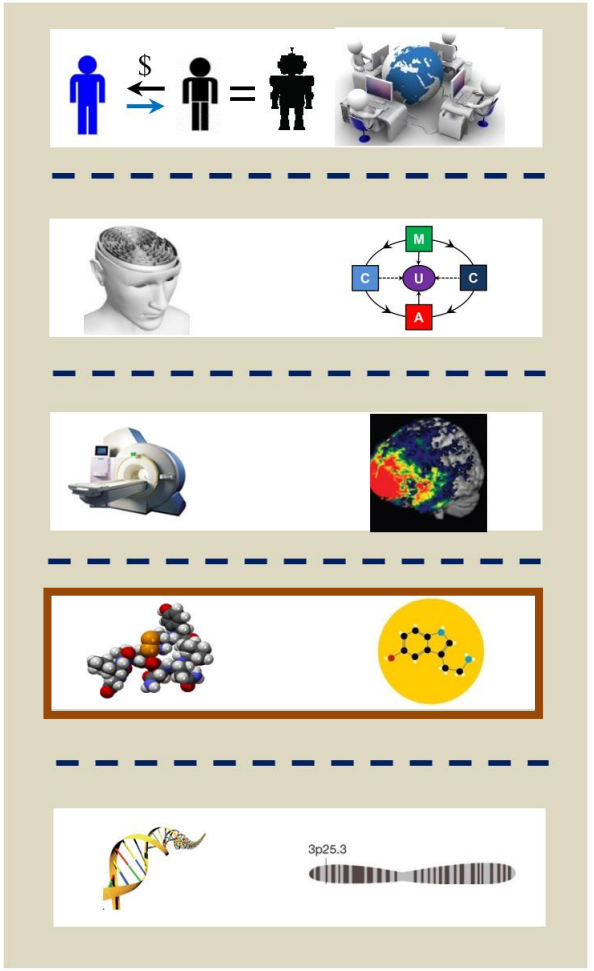
ALE: 30 papers, 10,000 permutations, cluster forming threshold of P < 0.001, k>90



Bellucci et al., 2016

NeuroPsychoEconomic Framework





Neuropeptide: Oxytocin

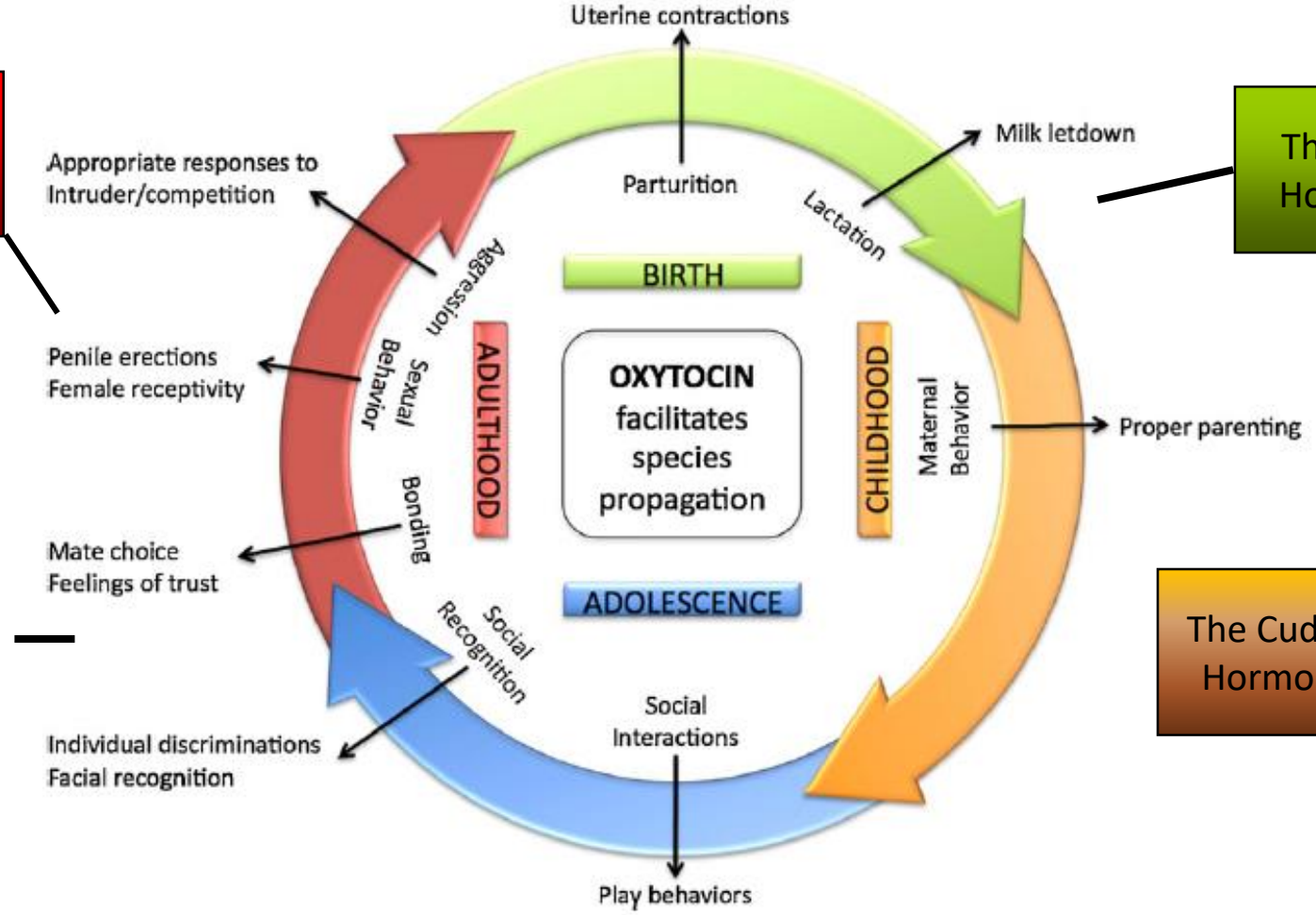
Oxytocin (quick birth)

The Love Hormone

The Birth Hormone

The Bonding Hormone

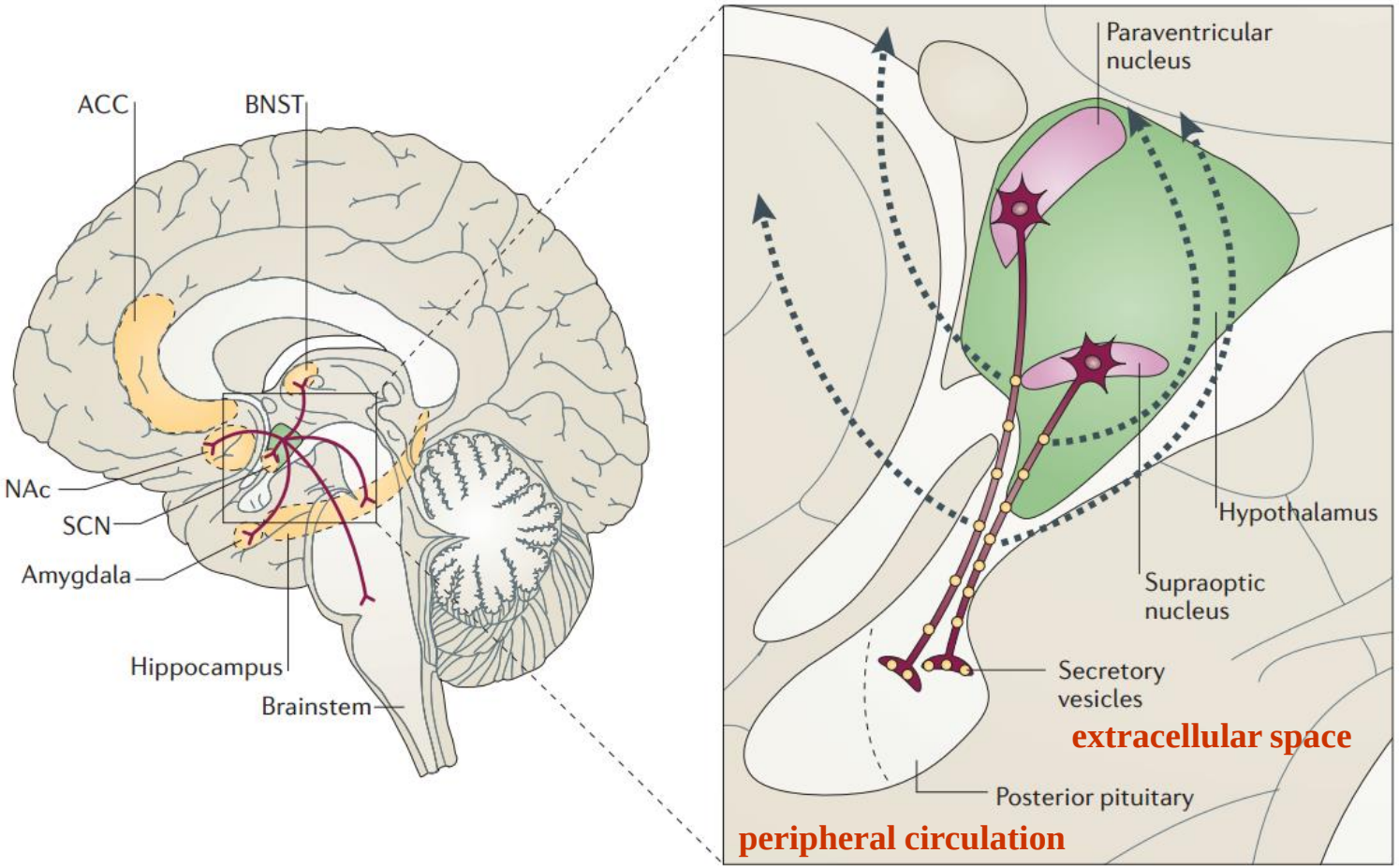
The Cuddle Hormone



Hormone & Neurotransmitter

Lee et al., 2009

Neuropeptide: Oxytocin



Meyer-Lindenberg et al., 2011

anterior cingulate cortex (ACC), suprachiasmatic nucleus (SCN), nucleus accumbens (NA), bed nucleus of stria terminalis (BNST)

Neuropeptide: Oxytocin

Interactionist Model

- Task difficulty
- Stimuli valence
- Group status

Contextual/ situational factors

Exogenous OXT administration

Mechanism(s)

- Perceptual selectivity/ social salience
- Anxiety reduction
- Affiliative motivation

Social Effects

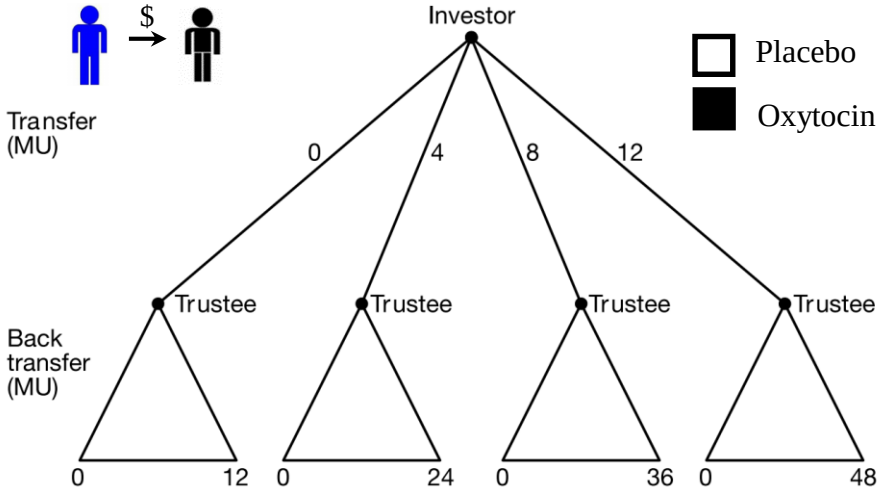
- Social Cognition
- Prosociality

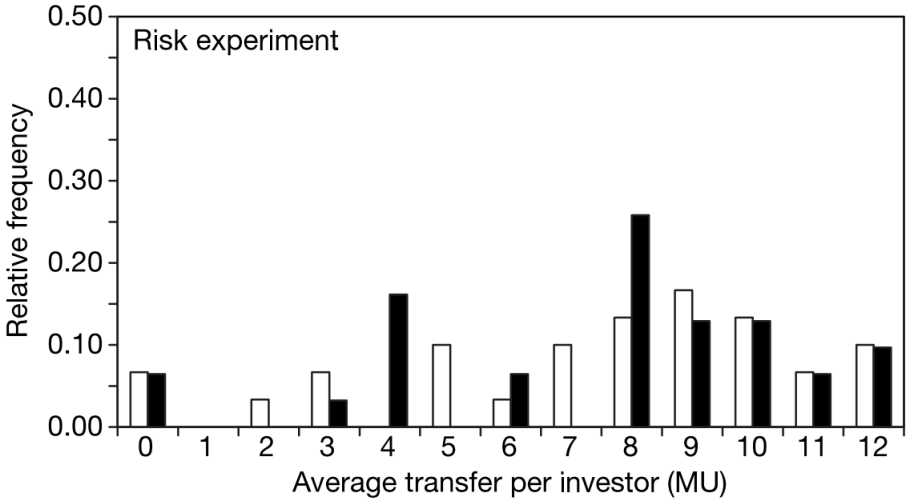
- Attachment style
- Mental Disorder
- OXTR Gene

Stable individual differences

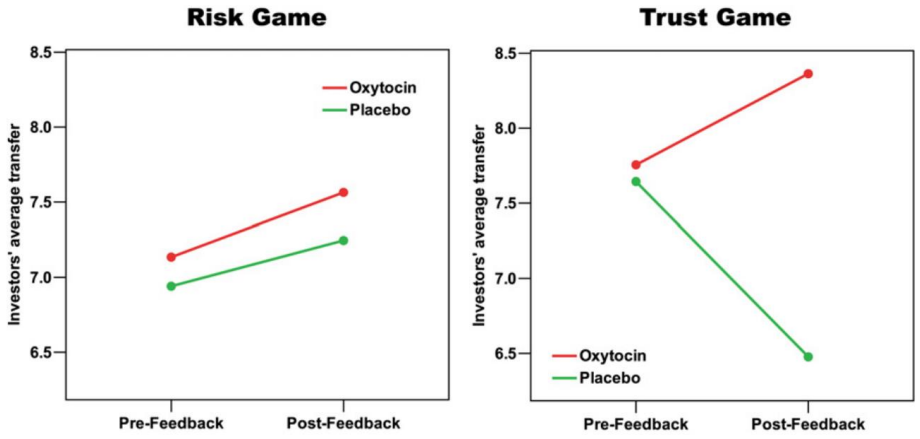
Bartz et al., 2011

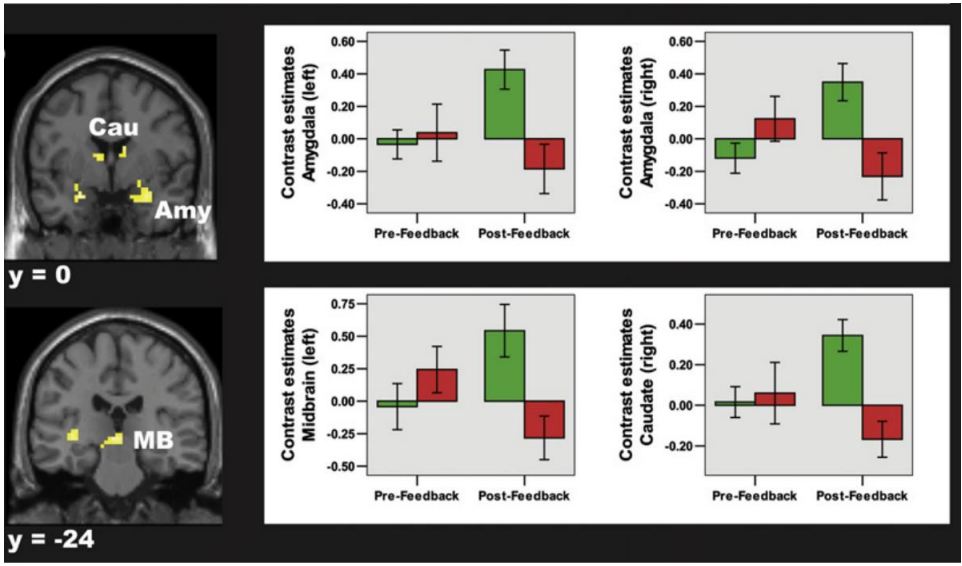
Neuropeptide: Oxytocin





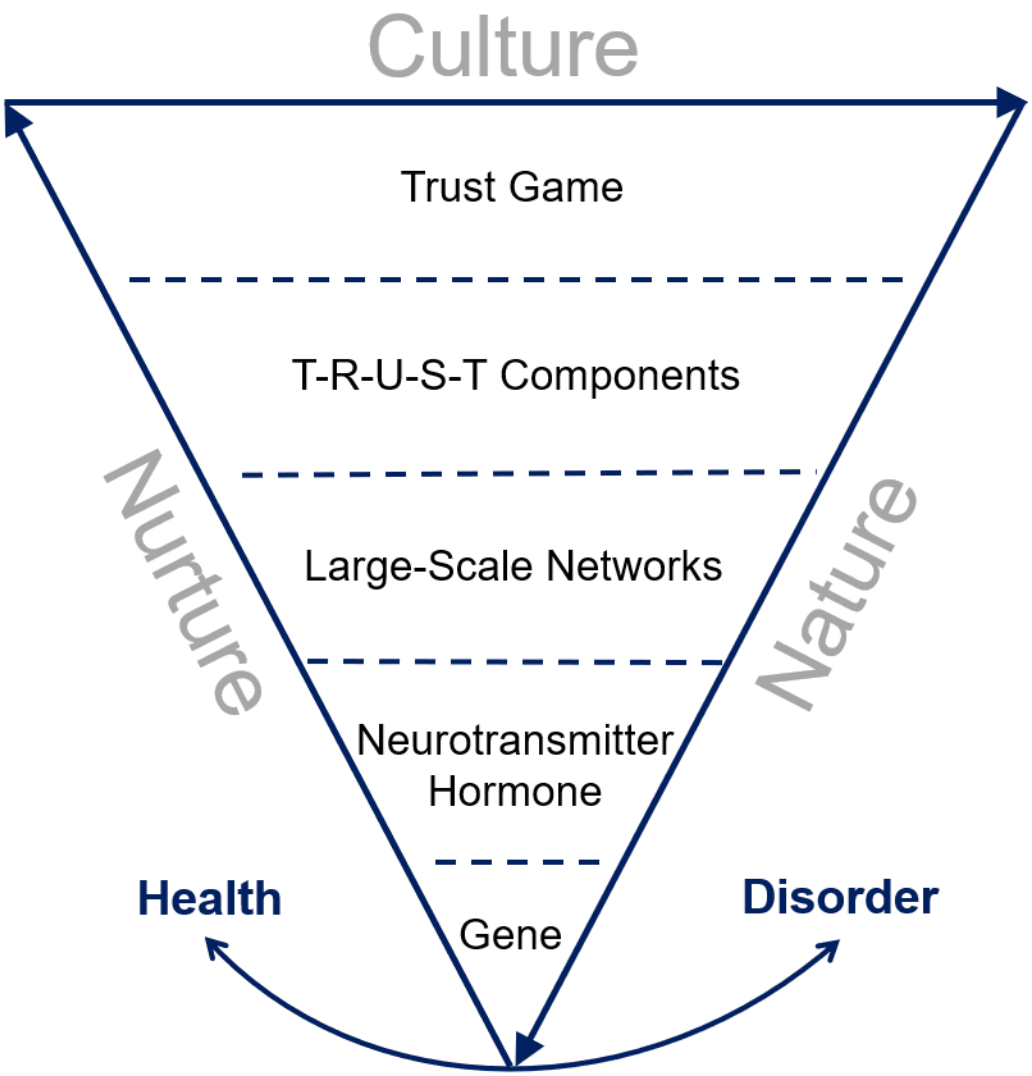
Kosfeld et al., 2005

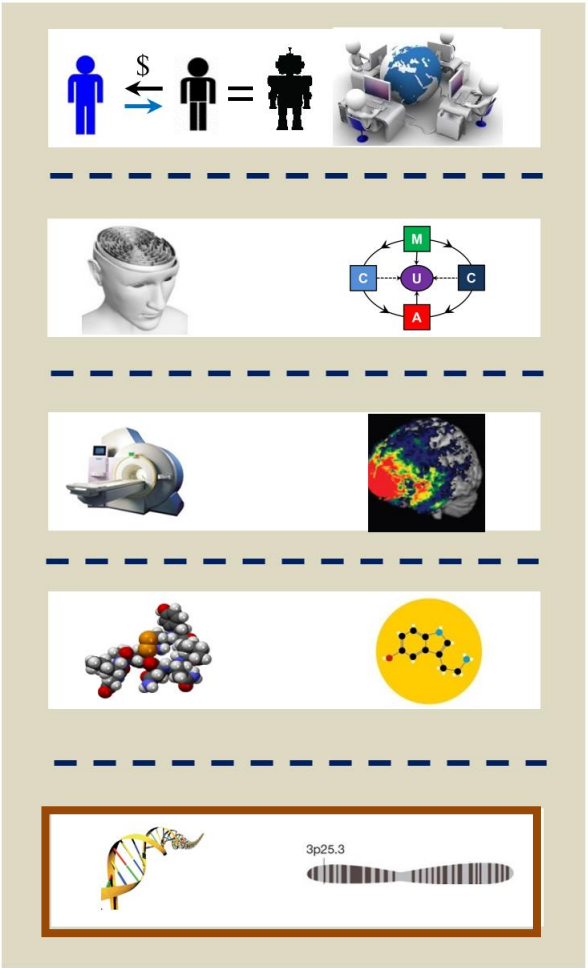




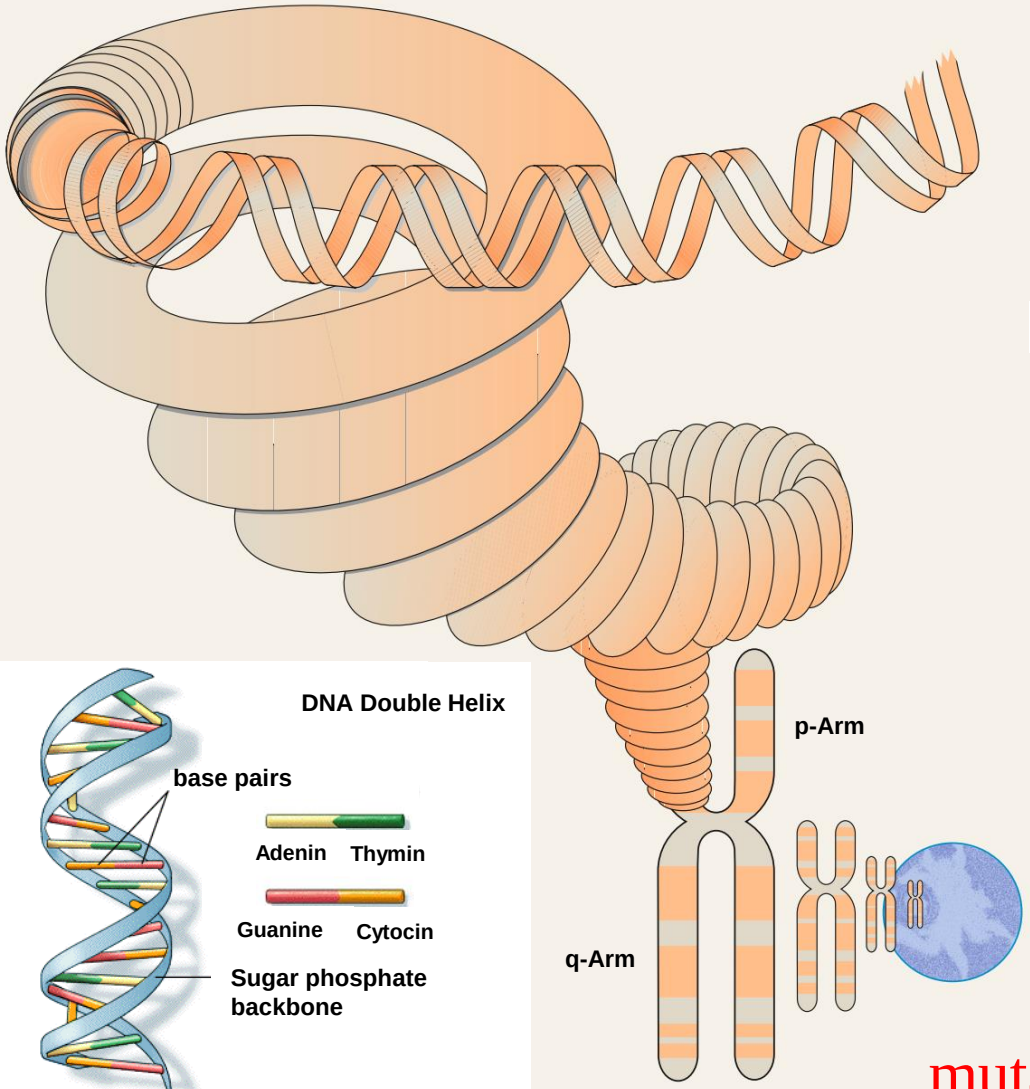
Baumgartner et al., 2008

NeuroPsychoEconomic Framework





Gene



Codon: 3 nucleotides (word) = 64 possibilities

Gen: Part of DNA molecule (many codons)

DNA: ~1.100 genes, chain of nucleotides

4 types: A, G, C, and TI

DNA = Deoxyribonucleic acid

Chromosomes: Protein + DNA (1.100 genes)

Nucleus: 23 chromosome pairs (30.000 genes)

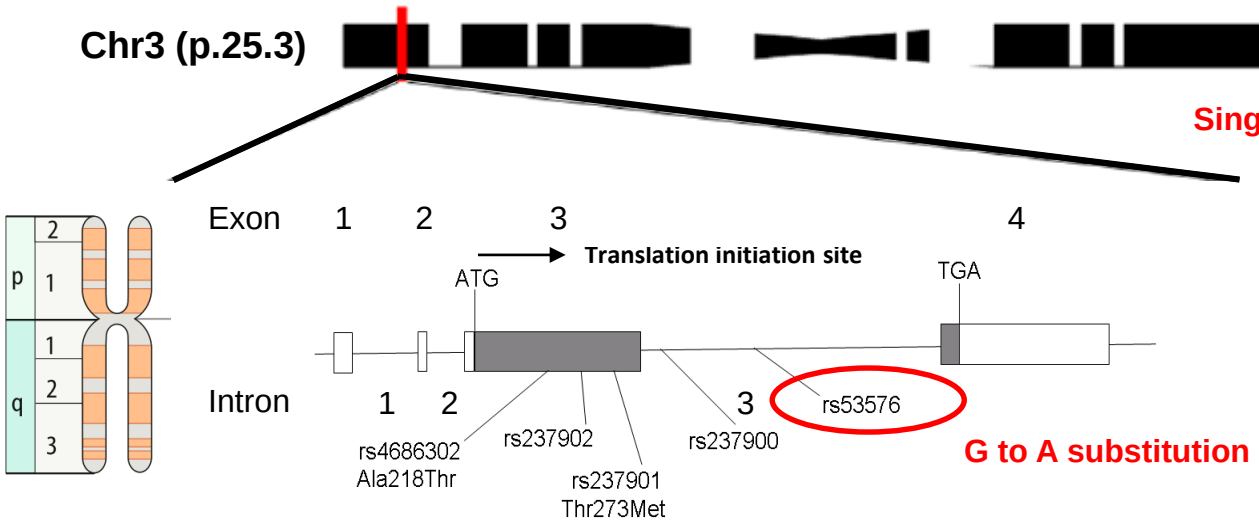
twin studies (10-20 %)

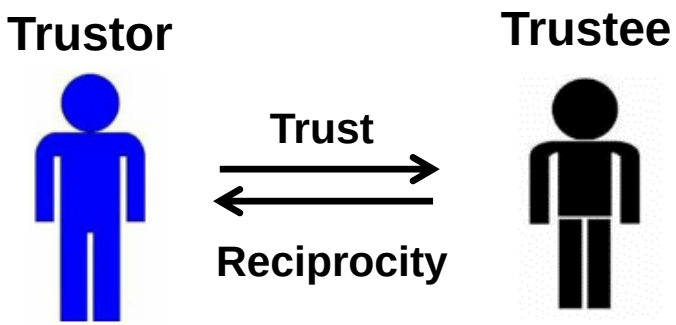
Cesarinie et al., 2008

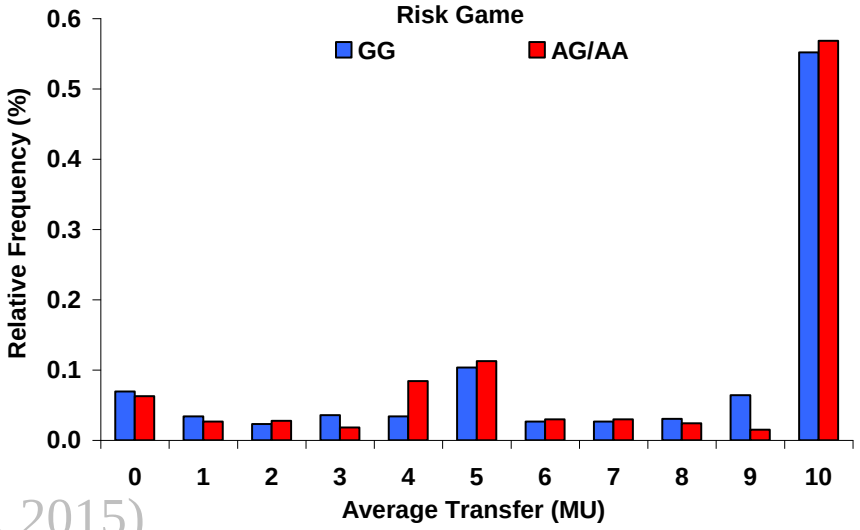
mutation ➡ polymorphism

Birbaumer & Schmitt 2010

Oxytocin-Receptor Gene



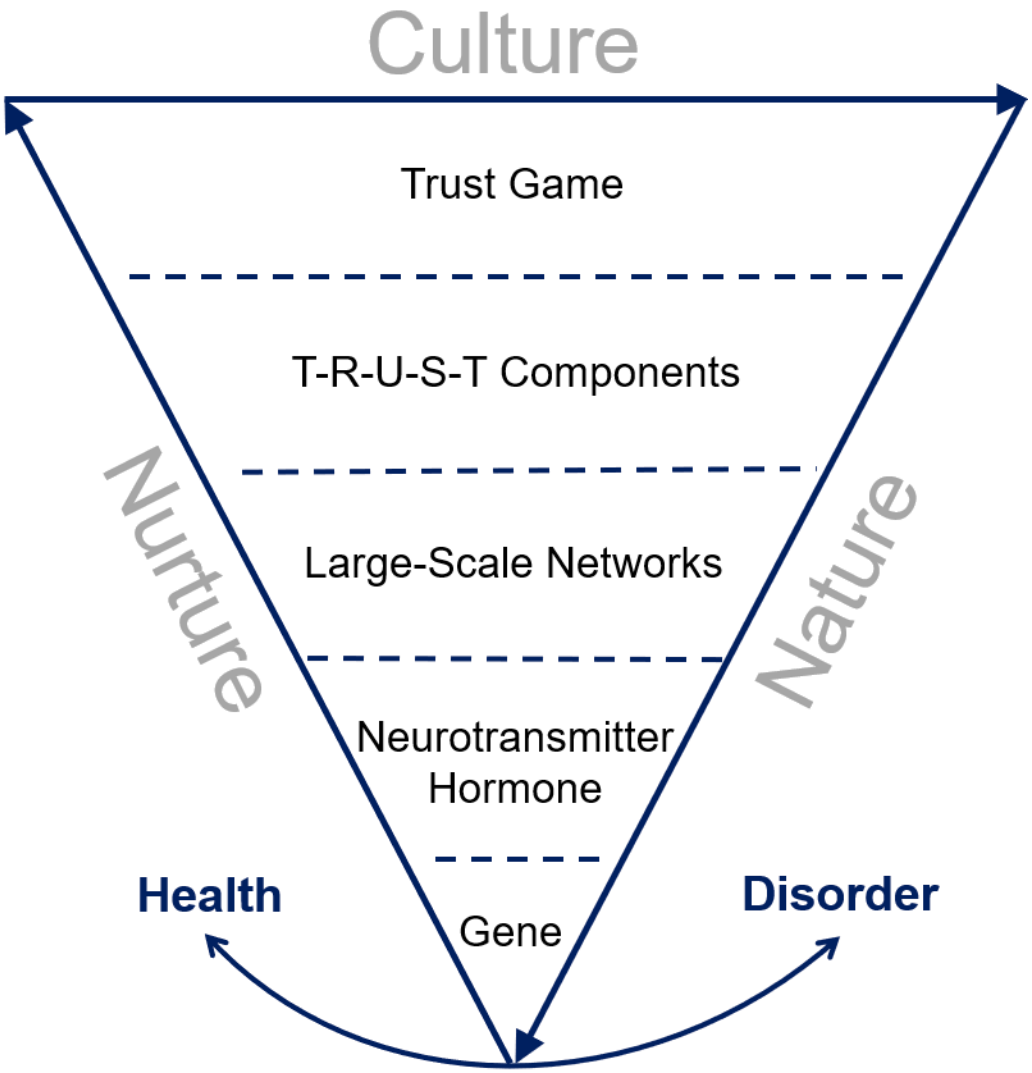


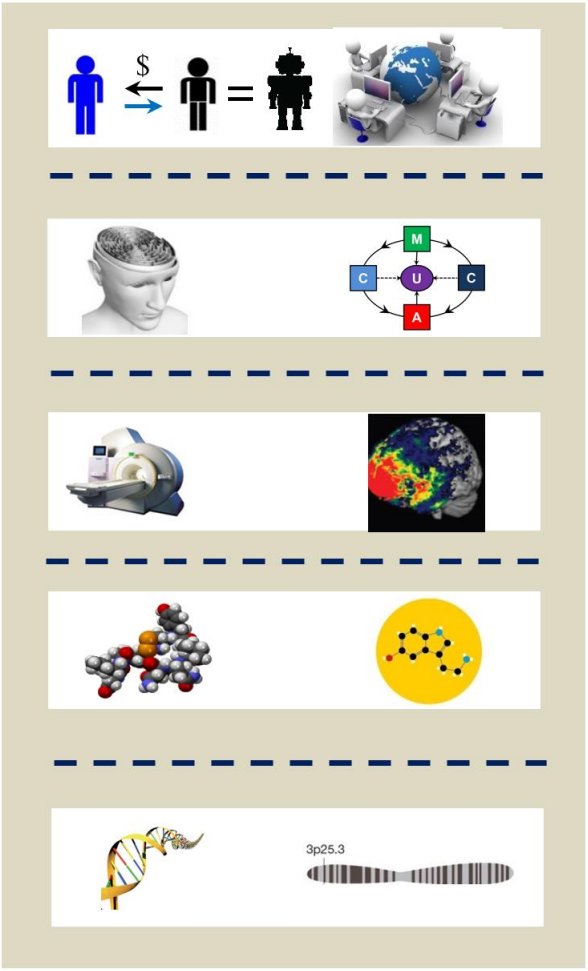


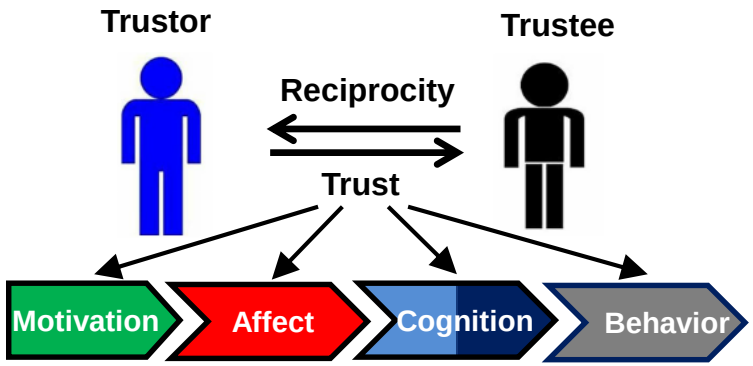
Robustness of OT Research (Nave et al., 2015)

Krueger et al., 2012

Summary: T-R-U-S-T Model



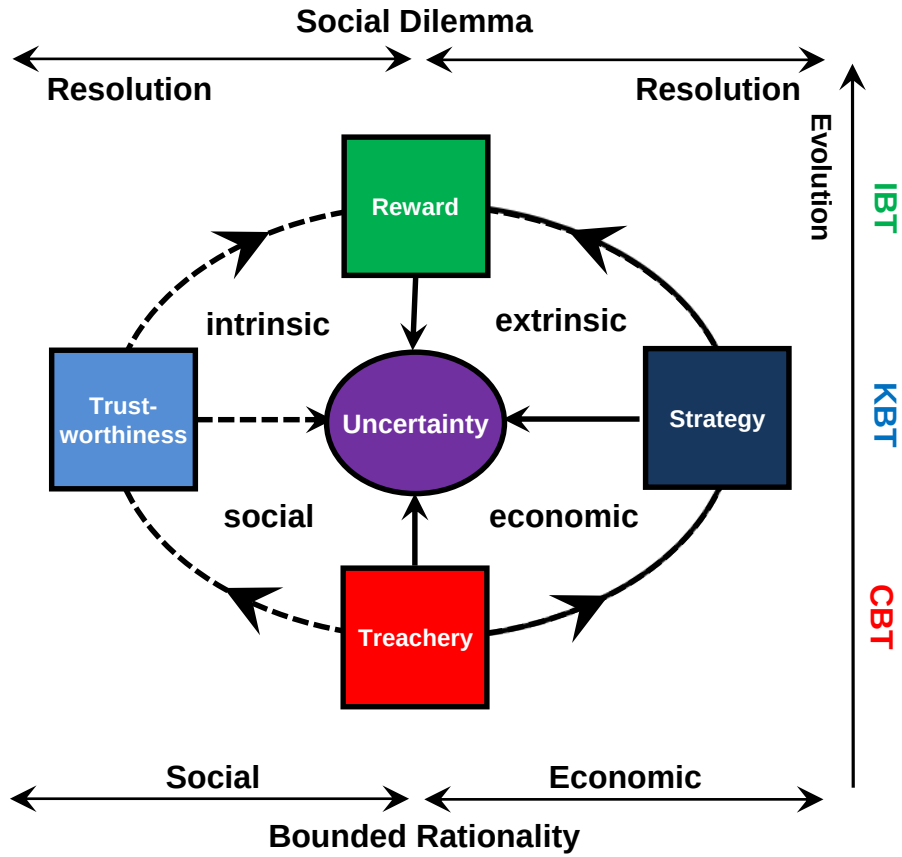


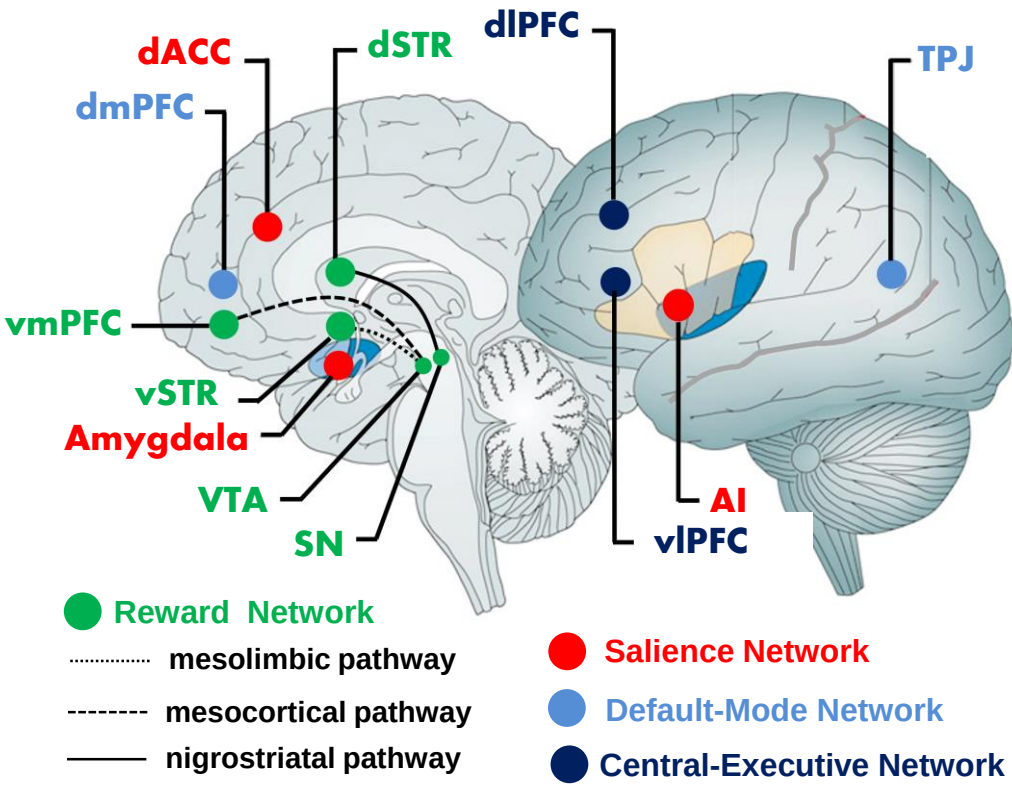


T-R-U-S-T Model

Formation **Resolution**

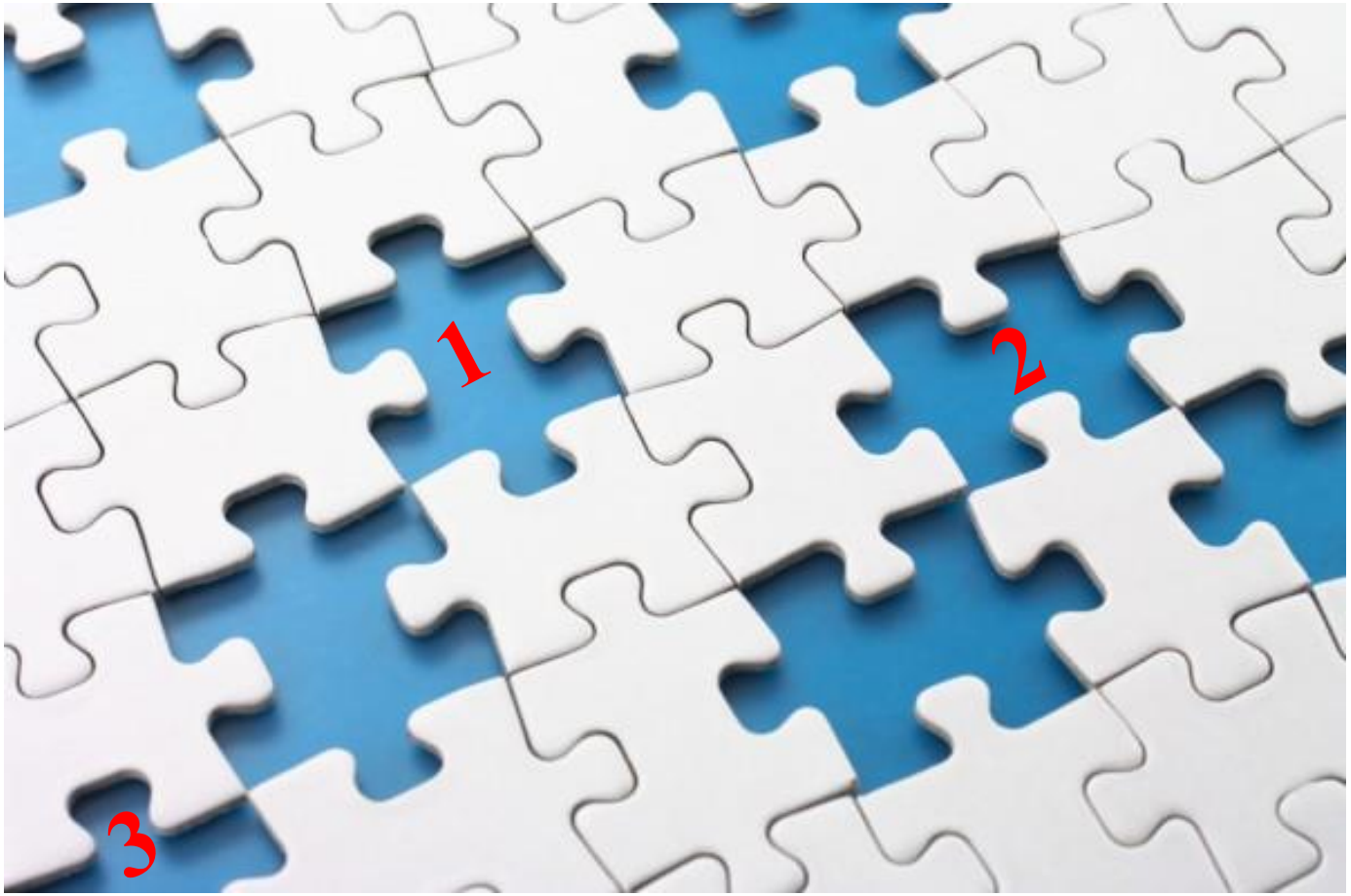
T – R –> U <– S – T





CBT, calculus-based trust; **KBT**, knowledge-based trust; **IBT**, identification-based trust; **VTA**, vental tegmentum area; **SN**; substantia nigra; **dSTR**, dorsal striatum; **vSTR**, ventral striatum; **vmPFC**, ventromedial prefrontal cortex; **AI**, anterior insula; **dACC**, dorsal anterior cingulate cortex; **TPJ**, temporoparietal junction; **dmPFC**; dorsomedial PFC; **dlPFC**, dorsolateral PFC; **vlPFC**, ventrolateral PFC

Outlook: What‘s next?



1 Paradigm

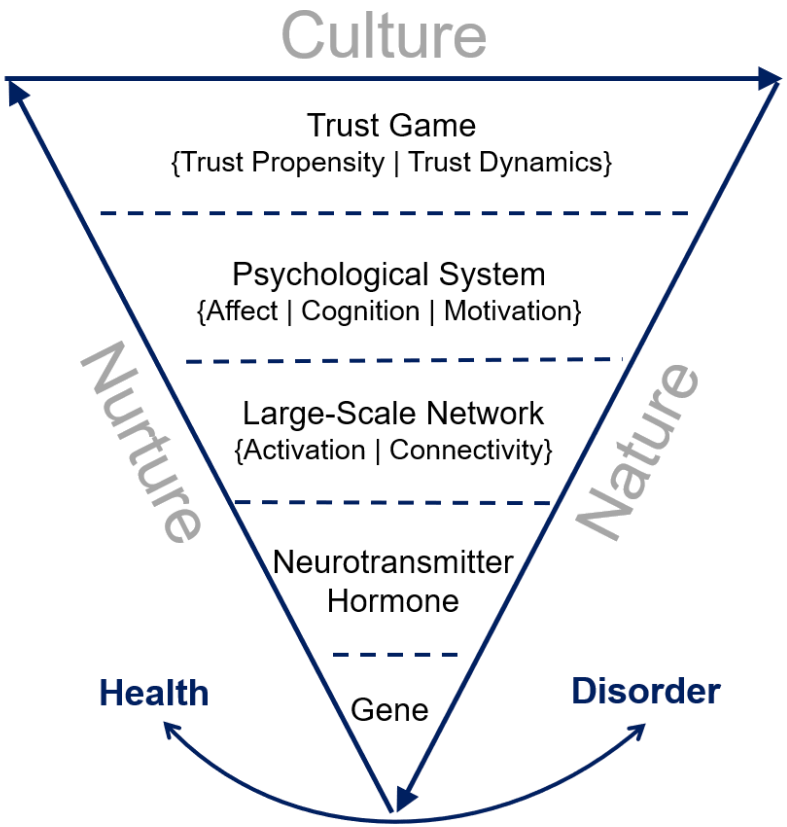
2 Mechanism

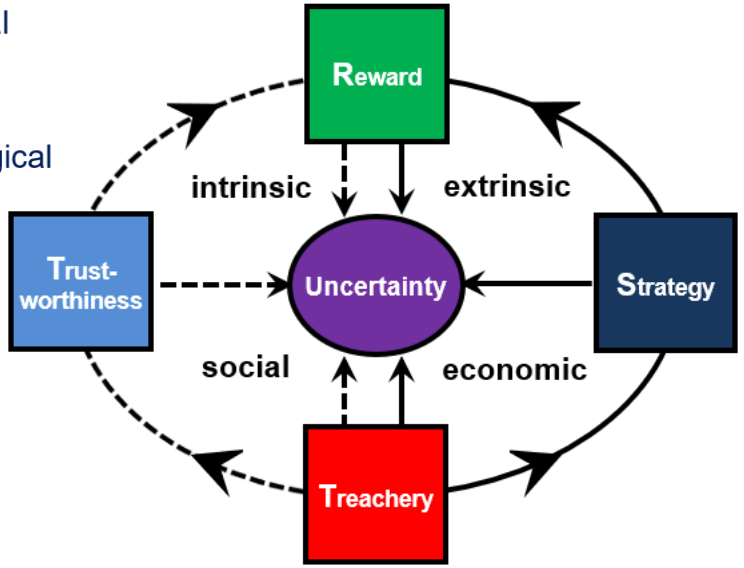
3 Application

T-R-U-S-T^3 Initiative

Toward a **R**efined **U**nderstanding of the **S**tudy of **T**rust (T-R-U-S-T)

Framework → T-R-U-S-T Model





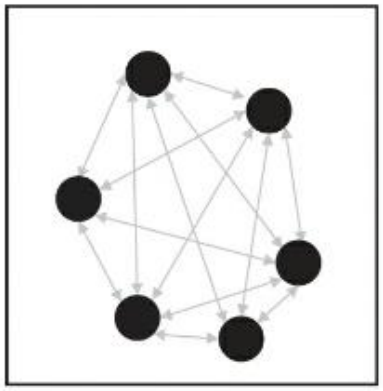
Transdisciplinary **R**esearch **U**nion for the **S**tudy of **T**rust (T-R-U-S-T)

Trust Architectures

Different architectures give rise to a trust trade-off, in which users give up some freedom to gain the benefits of trust.

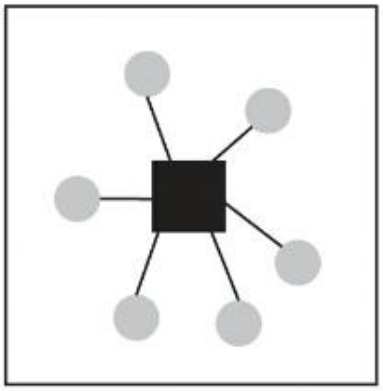
Black elments = trusted components

Peer-to-Peer



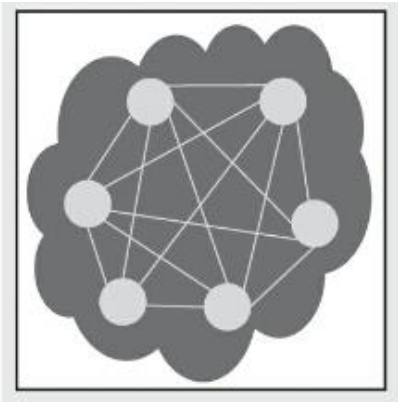
People must heed the norms of the community.

Intermediary



People cede control over personal data.

Blockchain



Promoting trust in a network without trusting any individual actor.

"Trustless Trust"

Power of Sanctions to Promote Trusts.

Centralized trust creates vulnerabilities and is not transitive.

Crypto-Economics

Blockchain — Too Trusted to Fail?

USING/ SELLING

Do I trust the companies accepting Bitcoins for purchases and the trading platforms?

Digitial currency exchange and all Companies, typically online buiseness today, that accept Bitcoin as currency

MOTIVATION

Do I trust dhe traditional monetary systems?

General government, central banks, financial institutions

Do I trust the algorithms and consensus mechanisms?

Potential Bitcoin User

Do I trust the legislative system and justice?

Political system and judicature

Digitial currency exchange which may be either brick-and-mortar business, and/or pure online businesses, etc.

Do I trust the trading platforms and others distribution channels?

BUYING

Bitcoin communitiy and miners, genéral government, central banks, financial institutions, Bitcoin startups, expersts, etc.

Do I trust the people and institutions that provide information on Bitcoin?

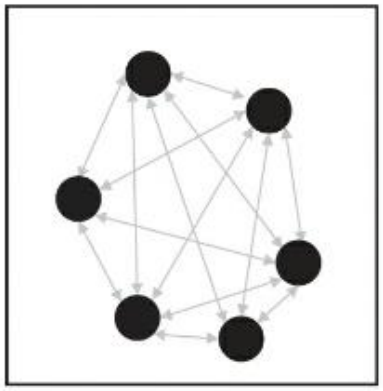
INFORMATION

Trust Architectures

Different architectures give rise to a trust trade-off, in which users give up some freedom to gain the benefits of trust.

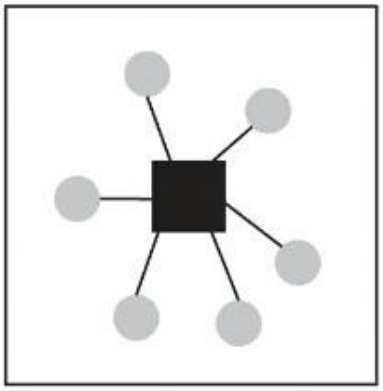
Black elments = trusted components

Peer-to-Peer



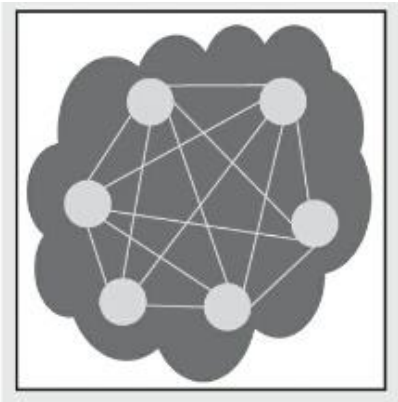
People must heed the norms of the community.

Intermediary



People cede control over personal data.

Blockchain



Promoting trust in a network without trusting any individual actor.

“Trustless Trust”

Power of Sanctions to Promote Trusts.

Centralized trust creates vulnerabilities and is not transitive.

Crypto-Economics => Crypto-Psycho-Economics

Towards a model of trust drawn from neuroscience, psychology, and economics:

Implications for the New and Next Generation Computing Environments

Frank Krueger, Ph.D.

Social Cognition and Interaction: Functional Imaging (SCI:FI) Lab

www.scifi-lab.com

FKrueger@gmu.de

SCI:FI LABORATORY

CROSSING Conference 2019
Darmstadt 9|10|19