What is Trust ?

Trust is *a* ***process*** (e.g., mental, physiological, cognitive, and behavioral) through which ***an agent*** *(e.g., human, machine, organization, or collective) evaluates* another system and regulates how to interact with it in order to achieve **a goal** (e.g., safety, efficiency, reliability, collaboration).

Aspect	Human Trust	General Trust	Human Trust in AI
Process	Mental, physiological (e.g., emotions, intuition, bodily reactions)	Information-processing and behavioral (e.g., reasoning, decision-making, observable actions)	Mental, physiological, and cognitive processes influenced by AI interaction
Agent (who trusts?)	Human individual	Generic agent (e.g., human, machine, organization, collective, institution, country, humanity)	Human (individual or collective) interacting with AI
Target (trusted entity)	Another system (human or machine)	Another system of any type	AI system, class of AI systems, or AI-embedded systems
Goal	Regulate interaction for safety, comfort, or satisfaction	Regulate interaction to achieve desired outcomes (e.g., reliability, efficiency, collaboration)	Regulate interaction with AI for fairness, reliability, transparency, and beneficial outcomes

Lukyanenko et al., *"Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities"* . https://doi.org/10.1007/s12525-021-00465-8

Trust

Trust is generally regarded as a **mental** (cognitif) and **physiological process** mechanism

- ***Reducing*** **uncertainty** means ***increasing*** the probability of a successful (e.g., safe, pleasant, satisfactory) interaction with entities in the environment.
- Trust has been evolution arily beneficial for humans/Be a prerequisite for any social interaction

Trust in technology → Trust in AI

→ **Use & Adoption (~AI that respects ethical principles** (absence of discriminatory bias, explainability, robustness, respect for privacy, etc.).

→ human autonomy, the right or the power to have control of own decision and choices, is one of the most common principles of ethical AI

- When **AI fails**, it can be due to:

  Bias in algorithms (quality of the data used),

  Discriminatory outcomes (e.g., COMPAS system used to aid judges: unfairly biased against African- Americans [*]),

  Unexpressed terms/assumptions => introduce bias.

[*] https://link.springer.com/article/10.1007/s10506-024-09389-8

Trust

- **HCI**: Human Computer Interaction: (The design + Psychological) → Impact users’ trust perception in AI system and their use behavior

- Such systems should be based on **solid engineering principles**, such as ***designing for failure***, ***having failsafe measures***, ***explicit maintenance protocols***, ***redundancy***, and ***design process transparency***

Trust

There is a **lack of consensus** in

on understanding the nature of ***trust***

Trust is a multidisciplinary research topic

(AI, Human-Computer Interaction, Computer Science, Sociology, Philosophy, Psychology, Marketing, Software Engineering, Information Systems, Medicine, Political Science, Economics, and Organizational Science).

Trust

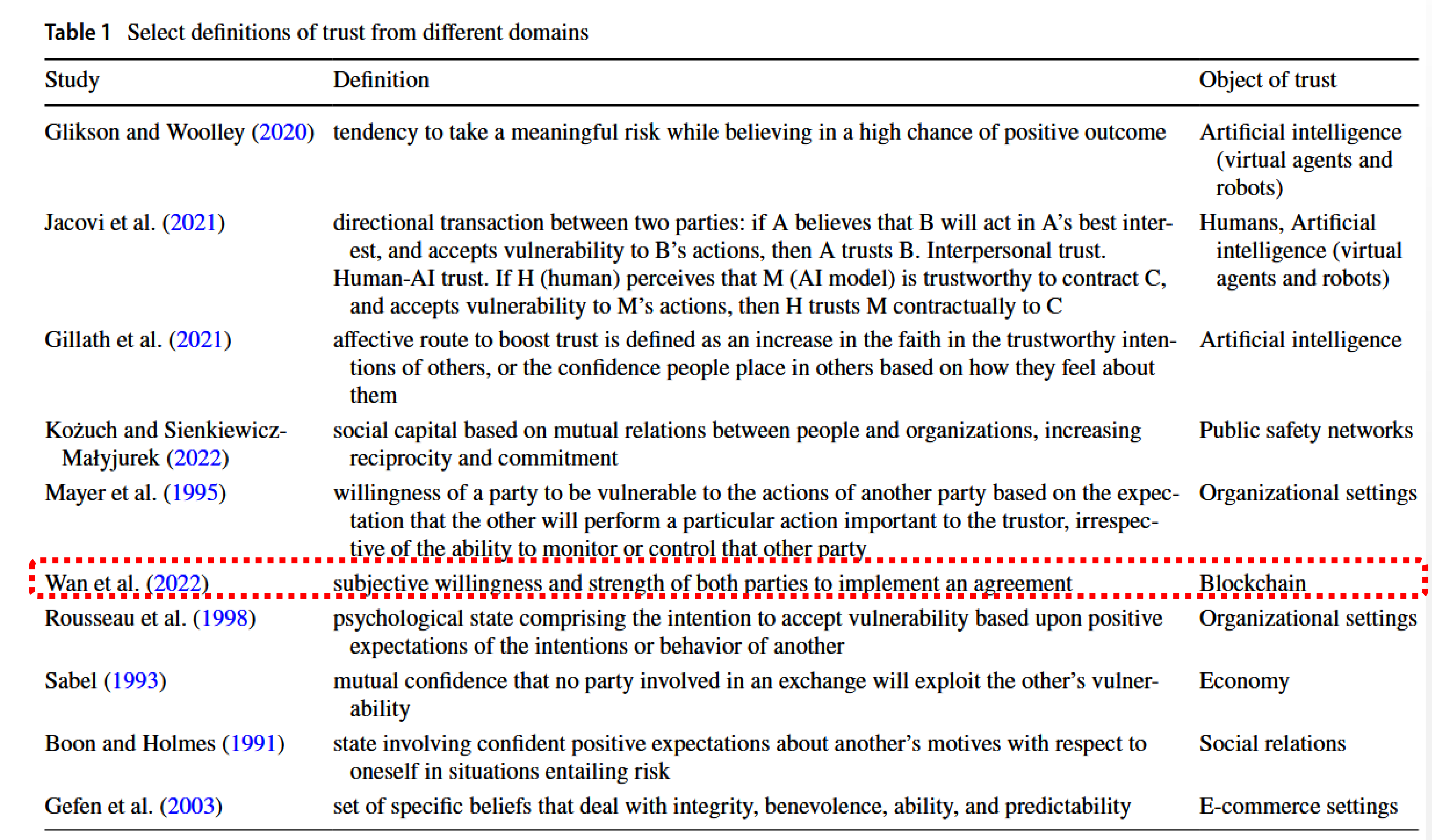
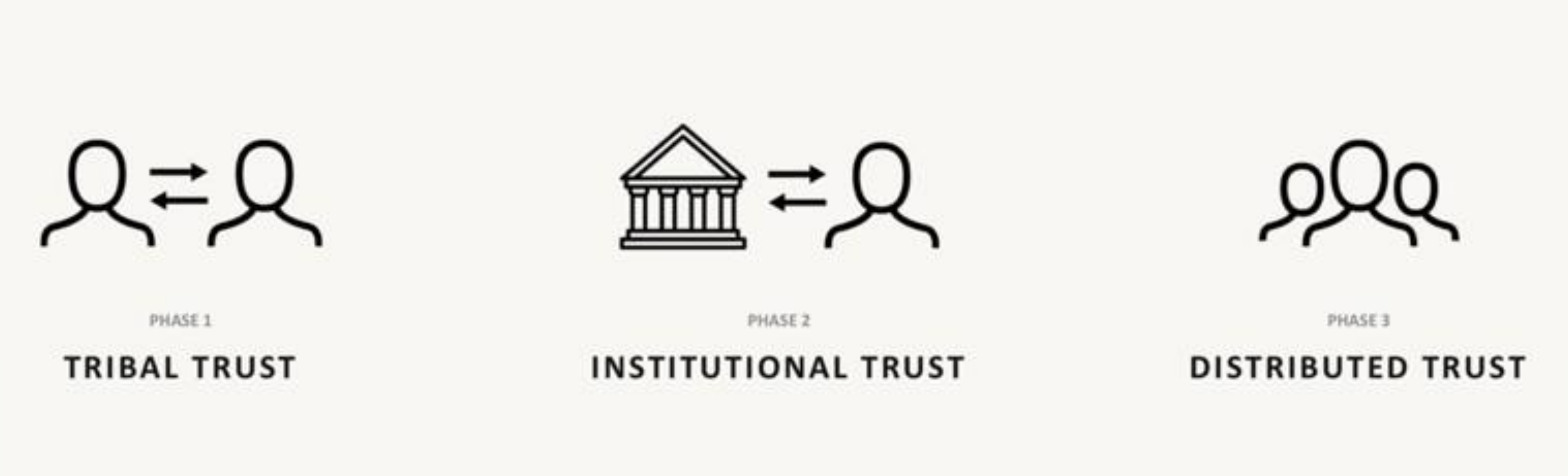


Table 1 Select definitions of trust from different domains

Study	Definition	Object of trust
Glikson and Woolley (2020)	tendency to take a meaningful risk while believing in a high chance of positive outcome	Artificial intelligence (virtual agents and robots)
Jacovi et al. (2021)	directional transaction between two parties: if A believes that B will act in A's best interest, and accepts vulnerability to B's actions, then A trusts B. Interpersonal trust. Human-AI trust. If H (human) perceives that M (AI model) is trustworthy to contract C, and accepts vulnerability to M's actions, then H trusts M contractually to C	Humans, Artificial intelligence (virtual agents and robots)
Gillath et al. (2021)	affective route to boost trust is defined as an increase in the faith in the trustworthy intentions of others, or the confidence people place in others based on how they feel about them	Artificial intelligence
Kożuch and Sienkiewicz-Małyjurek (2022)	social capital based on mutual relations between people and organizations, increasing reciprocity and commitment	Public safety networks
Mayer et al. (1995)	willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party	Organizational settings
Wan et al. (2022)	subjective willingness and strength of both parties to implement an agreement	Blockchain
Rousseau et al. (1998)	psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another	Organizational settings
Sabel (1993)	mutual confidence that no party involved in an exchange will exploit the other's vulnerability	Economy
Boon and Holmes (1991)	state involving confident positive expectations about another's motives with respect to oneself in situations entailing risk	Social relations
Gefen et al. (2003)	set of specific beliefs that deal with integrity, benevolence, ability, and predictability	E-commerce settings

Evolution Of Trust



PHASE 1

TRIBAL TRUST

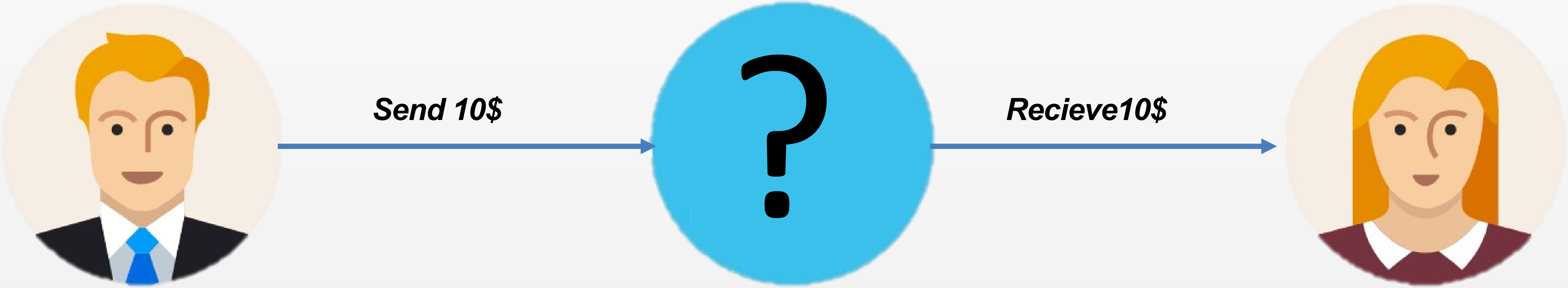
PHASE 2

INSTITUTIONAL TRUST

PHASE 3

DISTRIBUTED TRUST

Evolution Of Trust



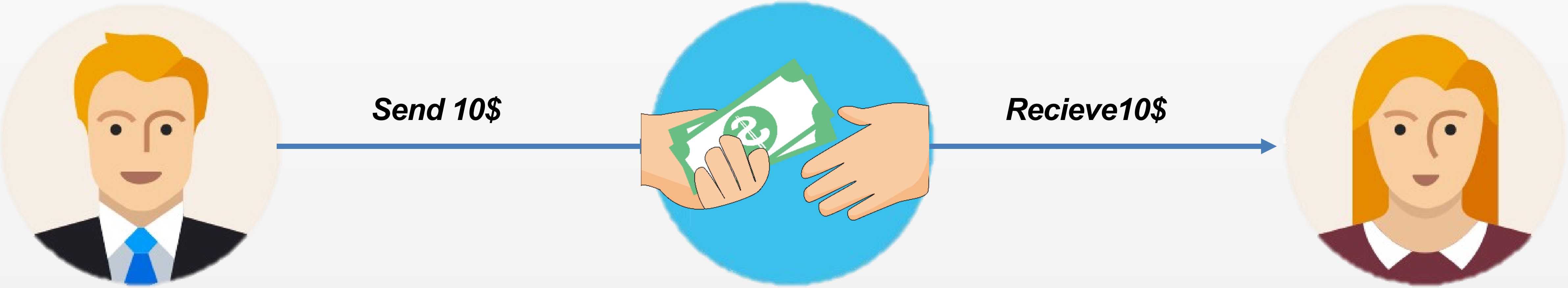
Send 10$

Recieve10$

Mohamed

Amina

Evolution Of Trust



Send 10$

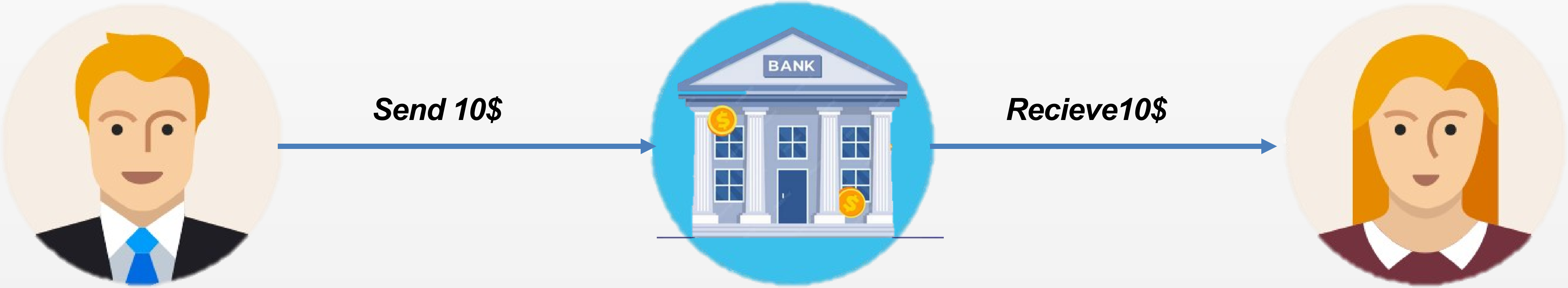
Recieve10$

Mohamed

Pay In Cash

Amina

Evolution Of Trust



Send 10$

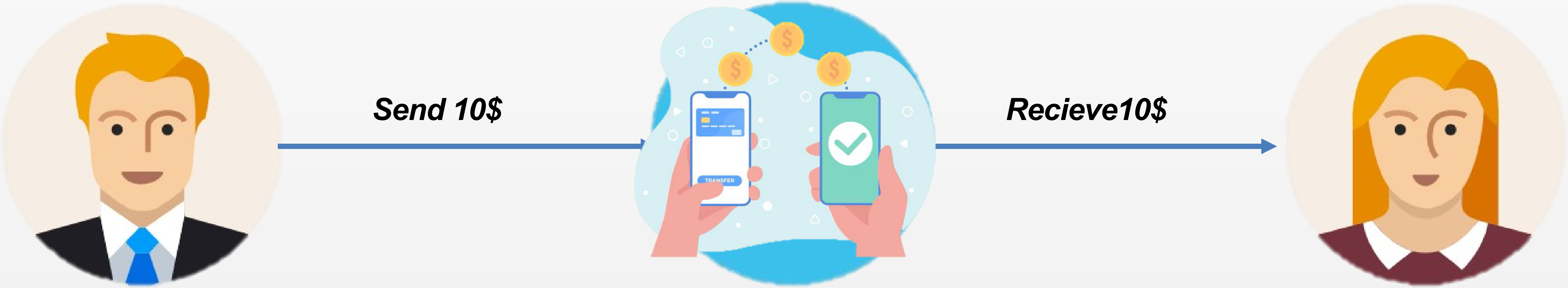
Recieve10$

Mohamed

Central Authority

Amina

Evolution Of Trust



Send 10$

Recieve10$

Mohamed

Third part application

Amina

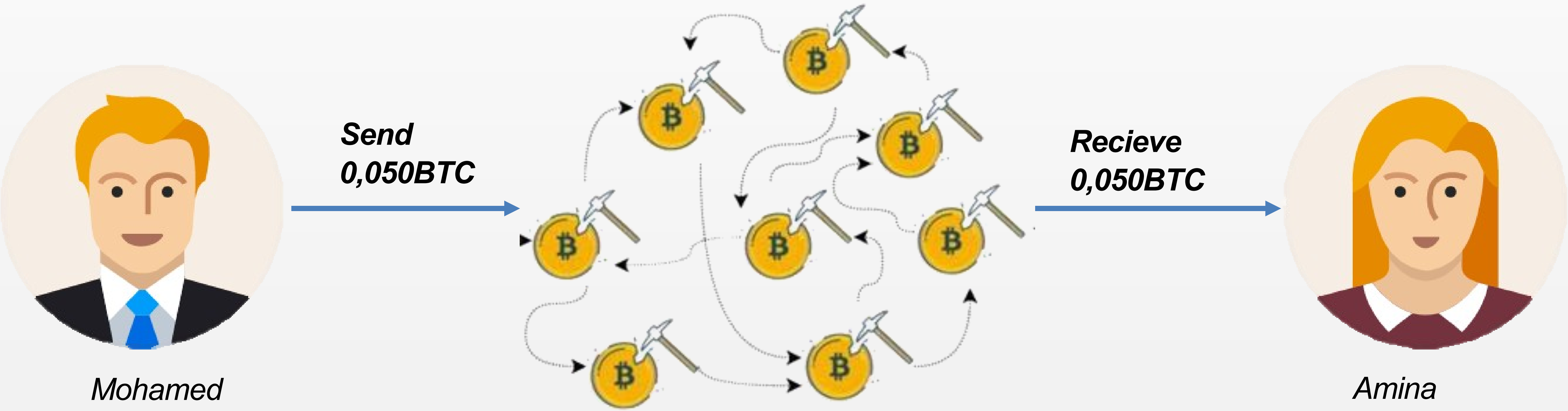
Two Problems :

- How to create a unique identifier without a central authority?

- How can I be sure that transactions were legitimately issued from an account?

Evolution Of Trust

Human success is based on flexible cooperation in large numbers. This requires **trust**



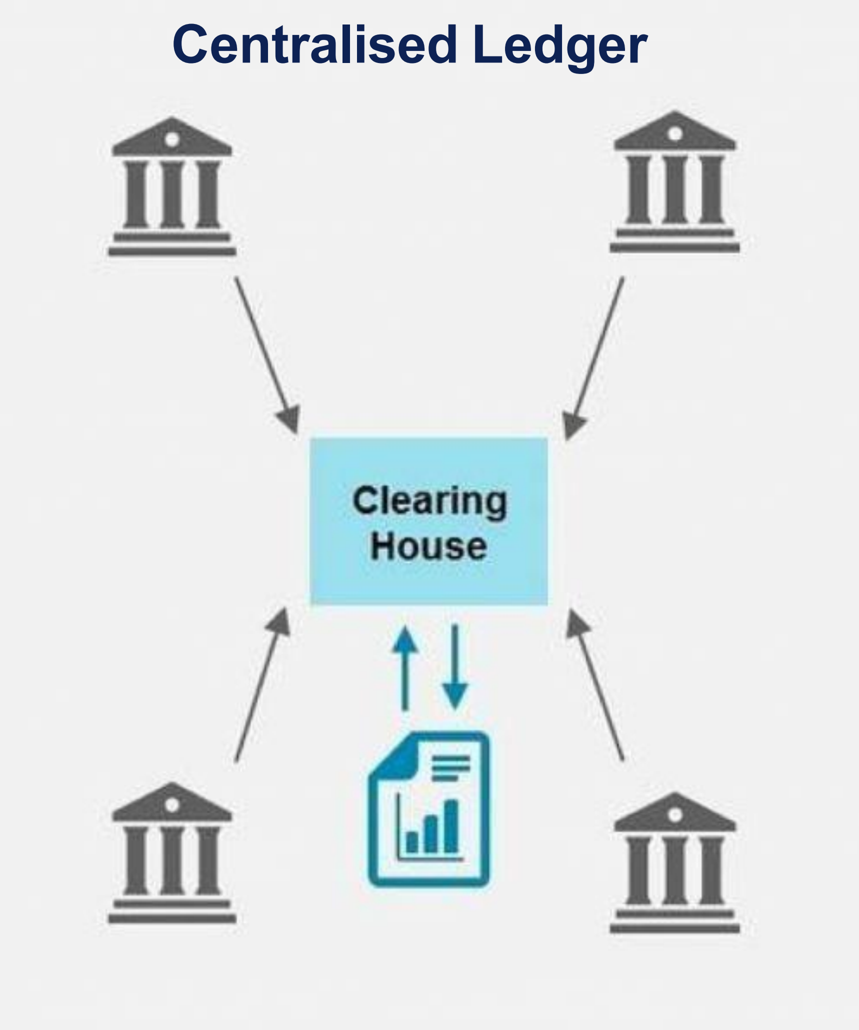
Send 0,050BTC

Recieve 0,050BTC

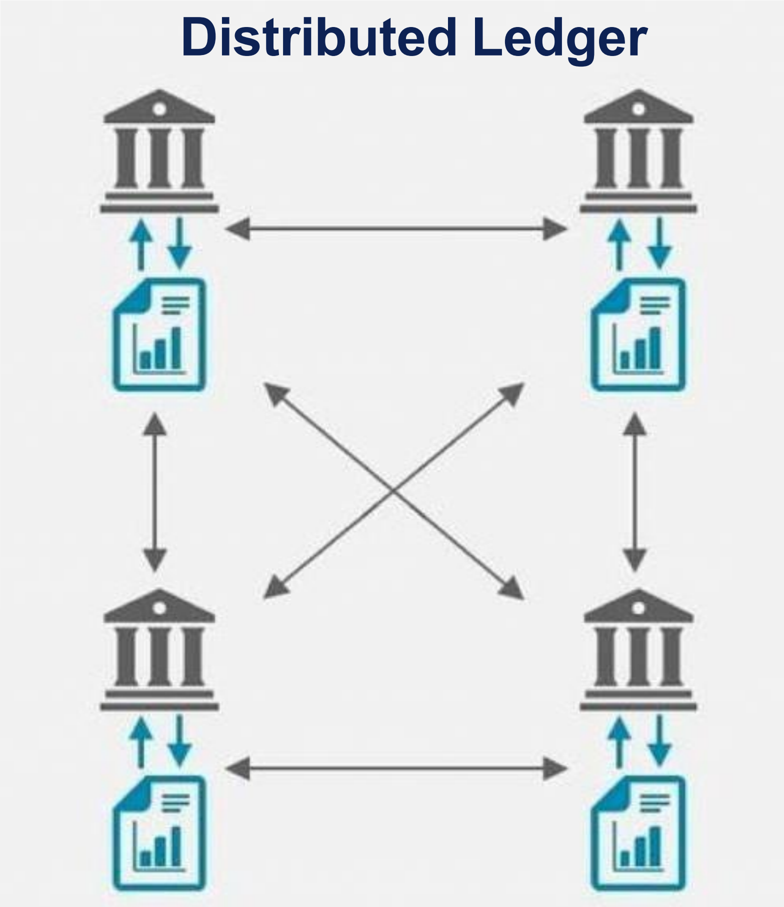
Mohamed

Amina

The main idea: Distribution and Replication



Distributed Ledger Technology (DLT)


Distributed Ledger