Money as an emergent coordination mechanism for economic networks: implications and limits for welfare economics

2026-08-04

Johannes Reich[1*] and Martin Härterich[1]

[1]SAP SE, Dietmar-Hopp-Allee 16, Walldorf, 69190, Germany.

*Corresponding author(s). E-mail(s): johannes.reich@sap.com;
Contributing authors: martin.haerterich@sap.com;

Abstract

A central obstacle in welfare economics is the assumed impossibility of interpersonal preference comparisons. This paper derives intersubjective comparability of valuations as a structural property of monetary exchange, rather than as a normative assumption or an implication of preferences under risk. We model money as a role-based coordination mechanism in which buyer and seller roles compose in two distinct ways: externally, through the exchange of goods and money, and internally, through the coordination of both roles within a single trading subject. Requiring these two modes of composition to be mutually consistent uniquely determines a linear utility representation — the quasilinear form standard in mechanism design – and under equal budgets and equal access to alternatives, this structure renders individual valuations intersubjectively comparable. We term these joint requirements the 'comparable valuation condition'. The framework clarifies both the reach and the limits of monetary comparison: willingness to pay becomes a valid welfare signal precisely when the condition holds, and breaks down in characterizable ways — due to budget inequality, non-substitutable goods, or preference distortion — when it fails. The same mechanism that enables decentralized coordination and organizational scaling also creates structural possibilities for exploitation and preference manipulation, explaining the ambivalent character of money. Finally, we argue that using preferences as a model of free decisions – which our derivation requires – forecloses their direct interpretation as welfare indicators. This tension, rather than being a limitation, opens a more honest and productive agenda for welfare economics.

Keywords: intersubjective comparison, preferences, money model, utility, free decisions

JEL Classification: D46 , D51 , D61 , D63 , D82 , B41 , C02 , D02

1 Introduction

There is a strong tradition in mainstream economics, that interpersonal comparisons of economically relevant preferences is meaningless, as the relevant measurement scale is of ordinal nature and it is assumed that no universal, interpersonal yardstick can be provided. In his historical perspective on this issue, Stavros Drakopoulos (Drakopoulos 1989, 2024) states that *"the majority of economists are prepared to declare that interpersonal comparisons [with respect to utility] are not possible."*

One popular quote is that of Lionel Robbins, who stated in (Robbins 1938) that *"every mind is inscrutable to every other mind and no common denominator of feeling is possible"* – and thereby repeated William S. Jevons' statement from 1871 (Jevons 1957, p. 14) almost verbatim. Kenneth Arrow wrote in his famous dissertation (Arrow 1951, p.9) that *"interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility."* Roger B. Myerson still stated in 1985 (Myerson 1985, p. 238) *"Interpersonal comparisons of utility cannot be given decision-theoretic significance. That is, there is no decision-theoretic meaning for a statement such as 'a movie gives me more utility than an opera gives you'."* And even Amartya Sen, who was *"much concerned with incorporating different ways of making interpersonal comparisons and their far-reaching consequences on what is permissible in welfare economics"* (Sen 2008), stated in 1999 (Sen 1999, p.68) *"If different persons have different preferences (reflected in, say, different demand functions), there is obviously no way of getting interpersonal comparisons from these diverse preferences"*.

If this were true, economics would face a severe dilemma. How should it say anything sensible about the distribution of scarce but exchangeable goods based on the preferences of the people, without being able to relate these preferences to one another?

John C. Harsanyi (Harsanyi 1955) and Amartya Sen (Sen 2017) made two different attempts to provide a solution to this dilemma. As we outline in Section 8 in more detail, John C. Harsanyi derives comparability from structural equivalence of individual and social preferences under risk, as expected-utility aggregation, thereby adding non-trivial assumptions about the form of preferences. Amartya Sen's concept of social welfare functionals accepts comparability as a normative input parameter, giving a flexible framework but not a derivation. Neither grounds comparability in the structure of exchange itself.

This is where our contribution steps in. Rather than treating money as a numéraire or utility proxy, as is standard in economics, we model it as an interactional mechanism that structures exchange relations and facilitates coordination among interacting subjects. Our novel approach is based on the insight that neither the economic subjects nor their interactions can be modeled independently. So we resort to one common building block that we call a 'role' for modeling both the subjects as well as their interactions, and to couple these roles in two different and complementary ways, namely externally (coupling to trade-protocol interactions, the edges of the graph) and internally (coupling to the economic subject as a trader, the nodes) within a virtually infinite graph of economic relations.

In Sections 3-6 we analyze the ensuing network of economic relations with monetary trades as a sequential exchange game, a stepwise interaction between buyer and seller roles – and show that additionally requiring subjects to coherently coordinate both roles as a trader uniquely pins down the required linear utility representation. This is Theorem 2 of the paper, and it is the formal engine driving the comparability results. Our central claim is that intersubjective comparability of valuations emerges as a property of role-based monetary interactions of subjects under the constraint of equal budgets and opportunities, which we call the 'comparable valuation condition', rather than as an assumption concerning subjects.

We essentially apply the measurement theory of David H. Krantz, R. Duncan Luce, Patrick Suppes and Amos Tversky (KLST) (Krantz et al. 1971), according to which the presence of a second dimension, like money, together with some independence assumptions leads to a possible interval scaling of an otherwise only ordinally scaled quantity.

Our results apply to material preferences, those guiding our decisions within a monetary exchange of substitutable goods – and the comparable valuation condition is demanding. When it fails, due to budget inequality, non-substitutable goods, or preference distortion, the comparability breaks down in precisely characterizable ways. This is more useful than either treating comparability as universally impossible or importing it as an ungrounded assumption.

Quite relevant for welfare economics, we argue in Section 7 that despite our proposed 'quasilinear' utility representation, our model retains the possibility to represent wealth effects by allowing our valuation to naturally depend on the available budget in a single trade interaction.

We also discuss how the ideas of mechanism design, where this utility model is ubiquitous (Börgers 2015, p.7), are used to develop empirical tests of the validity of the quasilinear money model.

We extend our model of a trade, in Section 10.3, to the important case where there is also a discretionary component on the seller side. The most important example here is the employment relationship as the 'trade of labor'.

In Section 9 we discuss how our approach relates to the traditional welfare-economic concepts of Pareto efficiency, willingness to pay, cost-benefit analysis and the measurement of utility difference with money metric functions.

In our last Section 10 we use the logical impossibility to accept our proposed money model and interpret preference-based utility as a direct measure for welfare at the same time as a mandate to propose a broaden scope of welfare economics.

We also extend our model of a trade with a discretionary component also on the seller's side for the important case of 'trading work against salary'. As a consequence, two important recursive effects on the preference relations become important that might influence salary determination.

Also, as the monetary layer cannot be cleaner than the preference layer beneath it, substitutability comes into our focus, or more precisely, the fact that with the money mechanism we construct a partial space of comparability where outside no coherent scalar measure of trade-off exists. In particular, the proposed model of the money mechanism explains the ambivalent character of money. With the same function that enables social cooperation, decentralized coordination, and organizational scalability, the money mechanism creates the structural possibilities for exploitation, dependency, and preference distortion. These two sides of the coin suggest that monetary mechanisms must be embedded within carefully designed institutional contexts.

And finally we stress the contribution of our model to our concept of truly free decisions, which can be diminished due to external coercion and internal manipulation of our autoregulation system, for example by drugs or rewards. We introduce two well-known complementary properties of the linear utility representation as evidence for the tension of the money concept, which we term 'Athena-Invariance' and 'Hermes-Desire'. If money worked according to our model, then by Hermes-Desire, money is particularly well suited as a reward in the psychological sense, tainting our preferences by shifting our motivations towards the outside. And by Athena-Invariance we perhaps have a sensitive schema to detect preference tainting, as money should not influence our preference relation of material goods. As truly free decisions do play such an important role for economics as such and for us as human beings as well, it would be a decent goal for welfare economics to investigate this subject further.

In summary, Sections 2-6 build the formal core, Sections 7-9 connect it to existing literature, and in Section 10 we try to suggest some starting points for broadening the perspective of welfare economics

2 A model of the economically acting subject

The economic interactions we wish to model have a sequential, stepwise structure familiar from extensive-form game theory. A trade between buyer and seller proceeds through a fixed sequence of moves: the buyer places an order; the seller accepts or rejects; on acceptance, the seller delivers the good; the buyer pays. Each party has one or more free decisions (to order or not; to accept or reject) and one or more enforced decisions (to pay upon delivery; to deliver upon acceptance).

This is the structure of a two-player sequential game with a predetermined protocol. Our formal model captures this using I/O-automata — finite-state machines that are a standard tool for modeling protocols with state, sequential steps, and nondeterminism resolved by decisions. The reader familiar with extensive-form games will find the concepts directly analogous: states correspond to game nodes, transitions correspond to moves, and the free decisions at nondeterministic transitions correspond to the choices available to each player at their decision nodes.

2.1 Interactions and coordination of the economic subject

We model the economic world as a virtually infinitely scalable network of economically interacting subjects, a section of which we show in Fig. 1. Thus, we propose to model this network as a graph, where the nodes are the subjects, and the edges are the interactions.

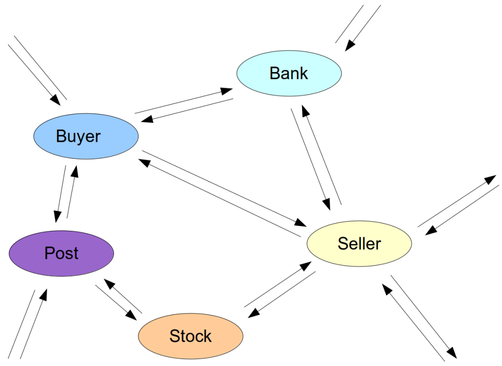

Fig. 1 A section of the infinitely scalable network of economically relevant interactions.

Beyond traditional network theory in economics or organizational psychology (e.g. Borgatti and Foster 2003), our model takes into account that the nodes (economic subjects) and the edges (economic interactions) of this graph cannot be modeled independently: we cannot describe the interactions without referring to the participating subjects, nor can we describe the subjects with their behavior without referring to their interactions.

The key idea, taken from computer science (Reich 2022, 2023), is to use a projection of the subjects, which we call a 'role' inspired by Erving Goffman (Goffman 1959), as the common building block for composing both the edges as well as the nodes of this graph.

The roles of different subjects compose – or, equivalently, couple – 'externally' by the exchange of information, goods, or money, resulting in an *'interaction'*, the edges of the graph. In informatics, another term is *'protocol'* (Holzmann 1991; Desai et al. 2005).

The roles of a single subject couple 'internally' by what we call a *'coordination'*, resulting in a (more complete) economic subject, a node of the graph. In game-theoretic terms, this internal composition of roles corresponds to a player participating in multiple games simultaneously, with decisions in one game potentially constraining or determining decisions in another.

2.2 The role model in mathematical terms

We describe the role model and its internal and external composition in detail to make the decision-making concept in economic interactions clearer and thereby the

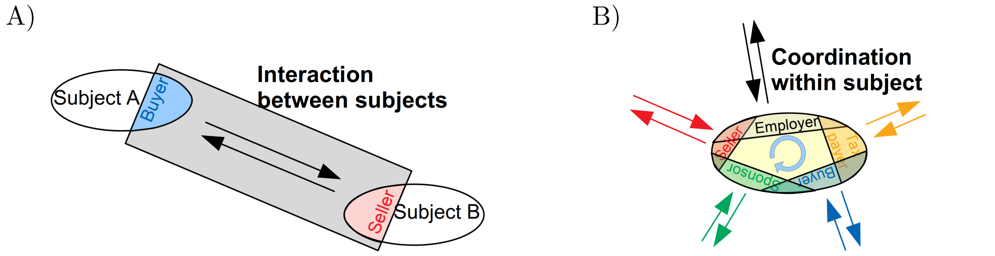

Fig. 2 The two different ways roles can compose correspond to two different perspectives. They either compose externally, creating an interaction (part A) or internally, creating a coordinating subject (part B).

relevance of the preference (and hence, utility) concept for modeling our free decisions later on.

We model a role as a discrete, stateful, and nondeterministic input/output transition system together with a success criterion as an input/output-automaton (I/O-A)[1].

Thus, an I/O-automaton consists of three sets of characters, I, O and (non-empty) Q, so-called *'alphabets'*, for input, output and internal state values, a start state q_0, a transition relation $\Delta \subseteq I^\epsilon \times O^\epsilon \times Q \times Q$, where $I^\epsilon = I \cup \{\epsilon\}$ and $O^\epsilon = O \cup \{\epsilon\}$ with ϵ the empty character. In case of finite interactions, the success criterion is the set of final states or *'outcomes'*[2].

Instead of $(i, o, p, q) \in \Delta$, we also write $p \overset{i/o}{\to} q$. In this model, a character might not carry any other semantics than being distinguishable from all other characters of its alphabet, being information in the sense of Claude E. Shannon (Shannon 1948), but it could also denote some 'real-world' entity like apples.

Why do not we use the extensive game structure instead of the I/O-automata, probably more familiar to an economically educated readership? Crucially, for our purposes we need a subject representation which explicitly allows for two different couplings – or compositions: An external coupling by character exchange, and an internal coupling by coordination. This cannot be represented in the extensive game structure, which abstracts away from the interactions and therefore cannot be used to distinguish the two composition operations. Both are structurally necessary for our central uniqueness result of Theorem 2 later on.

We can derive the extensive game form from the I/O-automata by equivalence class construction, in such a way that all state values that can be reached under the same decision are considered equivalent as introduced in the next section (Reich 2022).

[1] In the informatics literature, such an I/O-A is also called 'transducer', or, in the deterministic case, 'Mealy'-automaton.

[2] For infinite interactions, the appropriate success criterion generalizes the notion of a game's terminal payoff to sets of states visited infinitely often (for example as Muller acceptance) (Farwer 2001). In this paper, all relevant interactions are finite, so this generalization is not needed and the reader may safely ignore it.

2.3 Decisions of the economic subject

In general, the interactions will be nondeterministic, stateful, and asynchronous. They are nondeterministic and asynchronous because anything else would not scale, and stateful because otherwise nondeterminism would imply complete randomness.

In a nondeterministic transition relation, for a given input i and an internal state p, there could be several possible transitions that lead to different output characters o and new internal states q.

The transition relation can be determined with an additional alphabet D that is disjoint from $I \cup O \cup Q$ so that each nondeterministic transition is complemented by a unique character $d \in D$. We call these extra characters $d \in D$ 'decisions' and the complemented I/O-A a *'decision automaton'*.

As the transition relation of a decision automaton is deterministic per construction, the decision automaton determines a *'transition function'*[3]. In this model we can say that the nondeterminism of our interactions leaves us the freedom to determine our actions with our decisions.

We call a decision *'free'* if it is not determined by the success criterion of the interaction and *'enforced (by the success criterion)'* otherwise. Investigating how we make our free decisions constitutes the genuine economic part of this article. For this purpose we introduce a model of our mind in Section 4.

3 Interaction and coordination: the composition of buyer and seller roles to trade and trader

We now instantiate the general model with the two most fundamental economic roles: buyer and seller. Informally, the trade between them is a sequential game with the following stages: (1) the buyer decides whether to place an order at a given cost; (2) the seller decides whether to accept or reject; (3) if accepted, the seller delivers the good; (4) the buyer pays. The outcome for both is determined by their individual success criteria — completing the trade on acceptable terms. The formal model below makes this protocol precise, specifying the states, transitions, and coupling conditions that ensure the exchanged items and money amounts are the same for both parties.

To describe the roles of buyer and seller more intuitively (and more efficiently), we introduce a certain abstraction by summarizing the elements of the transition relation in equivalence classes, the formalism of which can be found in Appendix A. In essence we parameterize transition classes by a main state, some sent and received documents, a decision, and a condition.

3.1 Trade as interaction: external composition

In our trade interaction, the *external* coupling of two subjects in their buyer (B) and seller (S) roles is created by the exchange of money and a commodity a. The coupling condition is therefore the identity between B's cost and S's revenue which we call $price_a := cost_a^B = revenue_a^S$, together with the identity of the commodity for B and

[3] In the sense of (Reich 2022), the decision concept becomes a means to identify an interacting human as a system by a hypothetically assumed system function.

S and therefore an increment in the quantity n_a^B and a decrement in the quantity n_a^S in the handover step:

$$\begin{aligned}\text{External coupling:}\quad &(i)\ b^B \to b^B - price; \quad b^S \to b^S + price\\ &(ii)\ n_a^B \to n_a^B + 1; \qquad n_a^S \to n_a^S - 1\end{aligned} \tag{1}$$

In Fig. 3, we illustrate the external composition of a seller role S and a buyer role B to a simple trade interaction for a commodity a.

Buyer B has the state $(mod, n, b, val, cost)^B$ with mod^B being her mode state component, n^B being the number of desired commodities a, b^B being her budget, val^B being her valuation as the maximal amount she is willing to pay for a, and $cost^B$ being her cost she will have to pay to get a. We assume the initial state as $(\text{WISHFUL}, 0, b_0, val_a, \text{'undefined'})$. B proceeds as follows:

Being in mode WISHFUL B decides $@Order_p$ to buy a for some amount of money $cost^B$ with $b_0 \geq cost^B$ and signals this transition to S by sending an *Order* document with the necessary information to identify a and the amount $cost^B$. Then B awaits the delivery of the notification of whether S accepts the order or not. In case she receives a *Reject*, she transits back to mode WISHFUL. In case she receives an *Accept* she transits to mode CONFIRMED, waiting for a. If a arrives, n^B is increased by one and B transits to mode DELIVERED, where she has no alternative than to decide $@Pay$ to pay, i.e. to hand over the amount of money $cost^B$ to S that was determined in step 1 beforehand, by which her budget b^B shrinks by this amount while transiting into mode SETTLED. There B waits for the settlement notification *settled*. In case seller S is of the opinion that the trade had not been settled, she transits back to DELIVERED.

A bit simplified, B has to make two decisions[4]: First, she has to decide $@Order_p$ to buy a for given costs $cost^B$, and secondly, she has to decide $@Pay$ to pay. The first decision $@Order_p$ is free with respect to this interaction as it is not determined by the success criterion of the interaction, because WISHFUL is an allowed final state. In contrast, $@Pay$ is enforced by the success criterion, as DELIVERED is not an allowed final state and deciding $@Pay$ is the only way to proceed.

Now we turn to Seller S. He has the state $(mod, n, b, cost, revenue)^S$ where mod^S, n^S, b^S have a meaning equivalent to those for B and $cost^S$ is his cost to obtain a, for example through production, and $revenue^S$ being the amount of money he gets in exchange for a. S proceeds in the following steps:

Being in mode LISTENING S has produced a with costs $cost^S$ and waits for orders. If an order arrives, S transits to mode *ordered*. Now S has to decide whether he wants to accept the order or not. If he decides to reject, he transits back into mode LISTENING and sends back a *Reject*-document. If he decides to accept, S transits to the new mode CONFIRMED, indicating by sending an *Accept*. Being in mode CONFIRMED, S has no choice other than to decide to deliver the commodity a to B and to transit to DELIVERED to wait for the reception of the right amount of money. Upon receiving it, S increases his budget b^S by this amount and transits to its final state

[4]Effectively, $@Order_p$ is a class of mutually exclusive decisions that determine her cost.

SETTLED. Receiving too little money, he has to do some bookkeeping and sends back a *NotSettled*-Notification and waits for the rest.

To sum up the decisions of S: First, S has to alternatively decide whether to @*Accept* or @*Reject* and second, if he decides @*Accept*, he has to decide @*Deliver*. Again, the first decision is free with respect to the interaction as it is not determined by S's success criterion, while the second decision is enforced by the success criterion.

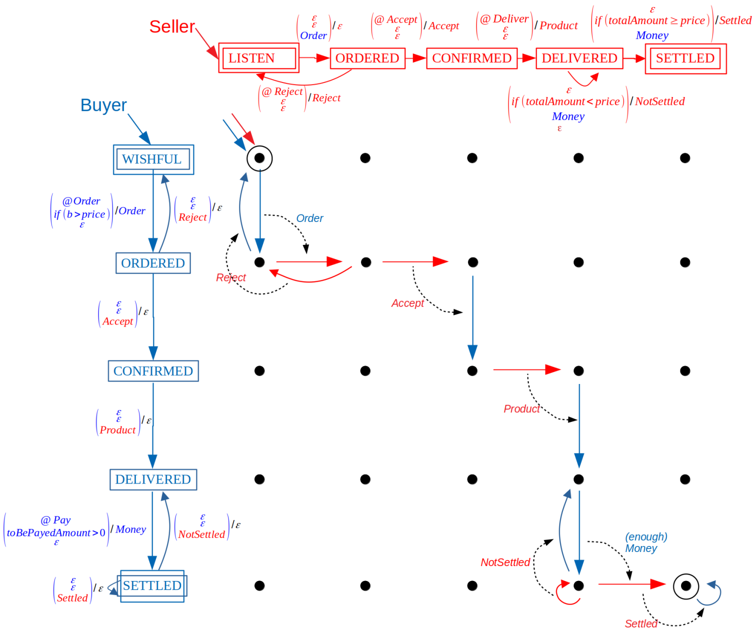

Fig. 3 A seller and a buyer engaged in a trade interaction as described in the text. All transitions of the buyer are lined up vertically and all transitions of the seller are lined up horizontally. Each dot represents a combined state of both buyer and seller. Each transition is labeled with $(decision, condition, incoming\ document)/outgoing\ document$. The external coupling results in a restricted reachable state space of the product automaton, illustrated by the 90°-arrows. A possible path of a complete trade is shown. Possible final states are indicated by an extra frame

To put it more formally, we model the resulting trade interaction *trade* as the result of the 'between subject', inhomogeneous[5] $comp^{external}$-operator such that it composes a B- and an S-role to a $trade = comp^{external}(B, S)$: the state of *trade* is the product state of B and S, the transition relation of *trade* is the restriction of the product transition relation, restricted by the character and commodity exchange. A trade is successful, if both roles are successful.

The model rests on two truthfulness assumptions, whose importance cannot be overemphasized, as they are easily overlooked:

[5] We name an operator 'homogeneous' if its domain is the same as its co-domain or any co-domain product (example $op : A \times A \to A$) and 'inhomogeneous' otherwise.

1. The exchanged items are truly identical for both roles. We represent this identity in our model by using the same identifier. Consider what happens if the seller believes he is receiving some euros but in fact he gets counterfeits, or if the buyer thinks she gets a tasty apple, but seller gives her a rotten one.
2. Both take their enforced decisions, that is, both adhere to the interaction rules that S delivers the ordered commodity and B pays the correct amount of money.

In reality, a verification of the truthfulness can be very complex, and therefore trust is an essential part of this kind of interaction.

We intentionally avoided the term 'price' in the description of the interaction. This illustrates that neither role needs to use or even to agree on this term during a trade.

However, introducing the term 'price' makes it much easier for us to understand what is going on during a trade, as it underlines the complementary meaning of the exchanged money: for the seller, the price is the money he receives, and for the buyer, the price is the money she has to provide.

As a final remark in this Section we want to point out that John Hicks (Hicks 1989) already emphasized the stepwise model of trading.

3.2 Trader as coordination: internal composition

Beyond external composition, roles belonging to the same subject can be composed in a second, internal way, which we call 'coordination'.

We illustrate a trader T internally coordinating her buyer role B_2 and seller role S_1 across two different trade interactions 1 and 2 in Fig. 4. Here, order acceptance within selling is granted only after acceptance of her own order is received within buying[6].

Coordinating roles means restricting the product transition space by coordination rules. First, the trader-state is *not* simply the product state of the roles; instead, we must declare which parts of the state are shared across roles. Secondly, we have to say which transitions become unavailable and, thirdly, which decisions may become dependent. Coordination requires all decisions to be named uniquely. The I/O-characters of the different, coordinated interactions need not be.

T has the state $((mode^S, mode^B), n, b, revenue, cost)^T$ and the coordination condition is:

1. Coordinating function of money: there is only one state for the budget $b^T = b^B = b^S$.
2. Identity of traded commodity: there is only one state for the possession of the commodity $n^T = n^B = n^S$.
3. Before T signals to agree to sell a as seller S in interaction 1, she must ensure that she acquires a in interaction 2 as B.

In game-theoretic terms, the trader is a player in two simultaneous sequential games, subject to a joint budget constraint. This is not merely a technical convenience:

[6] Other coordinations for T are possible, each with different chances and risks. For example, with enough budget it would be possible to order first, not knowing whether she will be successfully able to sell the produced commodity at a higher price than the purchase price. Or T could even order first without budget, hoping to be able to pay for the purchased commodity exclusively by the revenue it hopes to generate selling it. Or, in a mass case, T could use a mixed approach.

the requirement that a single subject can coherently occupy both the buyer role in one trade and the seller role in another is what forces the utility representations of both roles to be structurally compatible later on. We will see in Section 5 that this compatibility requirement uniquely pins down the linear form of both utility functions.

The following transition description of T should be read in conjunction with Fig. 4 and may be skipped by readers who find the figure self-explanatory. While in her mode (LISTEN, WISHFUL) she receives an $Order_1$ for a with price p_1 in her trade interaction 1 as S_1, which would incur a revenue $revenue = p_1$ and transits to (ORDERED, WISHFUL). She then has to decide $@Order$ to order as B_2 the same commodity at a price p_2 with $p_1 > p_2$, which would incur costs $cost = p_2$ and transit to (ORDERED, ORDERED) while sending the respective $Order_2$. If she instead decides $@Reject$ to reject the order, T has to transit back to (LISTEN, WISHFUL) while sending a $Reject_1$ in her trade interaction 1. Now, being in mode (ORDERED, ORDERED), T in her role B_2 receives either a $Reject_2$, requiring a backward transition to (ORDERED, WISHFUL), or an $Accept_2$, initiating the transition to (ORDERED, CONFIRMED). There, T must decide $@Accept$, sends an $Accept_1$ and transits to (CONFIRMED, CONFIRMED), awaiting the reception of commodity $Product_2$, which initiates her transition to (ORDERED, DELIVERED) and increases her commodity counter state to $n + 1$. Then, she has to decide $@Deliver$ to deliver the received commodity as $Product_1$, transiting to (DELIVERED, DELIVERED) and decreases its commodity counter n again by 1. There, she waits for the $Money_1$. If the received money suffices, that is, it was at least price p_2 then she transits to (SETTLED, DELIVERED), sends $Settled_1$ and increases her budget b by this amount. Then, she has to decide $@Pay$ to pay $Money_2$ to transit to (SETTLED, SETTLED) and reduce her budget b by this amount and finally wait for the last $Settled_2$-notification.

All other transitions of the unrestricted product automaton are eliminated by this coordination.

Again, more formally, we define a within-subject composition operator $comp^{internal}$ for a homogeneous composition such that we combine the two roles B and S of the same subject to a trader role $T = comp^{internal}(B, S)$. Part of $comp^{internal}$ are two projection functions[7] $proj^T_{B,S}\colon Q^T \cup I^T \cup O^T \cup D^T \to Q^T \cup I^T \cup O^T \cup D^T$ with $proj^T_B(Q^T \cup I^T \cup O^T \cup D_T) \subseteq Q^B \cup I^B \cup O^B \cup D^B$ and $proj^T_S(Q^T \cup I^T \cup O^T \cup D^T) \subseteq Q^S \cup I^S \cup O^S \cup D^S$. Δ^T fulfills the requirements that the two projections of the transition relation Δ^T of T result in the transition relations Δ^B and Δ^S and the success criterion of T is that both the success criteria of B and S can be fulfilled on the projected sets.

In the proposed coordination, T has to make 4 decisions, but only the first is still free. Namely, she freely decides either $@Reject$ to reject the received order $Order_1$ in interaction 1 or $@Order_{p_2}$ to order the commodity herself in interaction 2 for the price p_2. All other decisions, $@Accept$ to accept the order in interaction 1, $@Deliver$ to deliver the commodity in interaction 1 and $@Pay$ to pay the price $price_2$ in interaction 2, are enforced by the success criteria of both interactions and the coordination condition.

Interestingly, viewed in isolation, both B and S still have a free decision, but one of these free decisions is eliminated by this special coordination, namely S's $@Accept$.

[7] for a projection function h it holds that sequential application does not change the mapping, i.e. $h = h \circ h$

From an intuitive position, we would say that T has already decided to accept the order for internal processing with her decision $@Order_p$ to order the commodity herself in interaction 2 for price p_2. But this is invisible in interaction 1. We see that coordination changes the structure of decision-making.

We therefore speak of 'truly' free decisions of a subject only if it has the freedom to determine its behavior to achieve its objectives coordinating *all* of its roles, and – very important – it has the freedom to set its objectives. Otherwise, we would qualify decisions between several undesired outcomes as 'truly free', like a person sentenced to death who can still 'decide' to be shot or hanged.

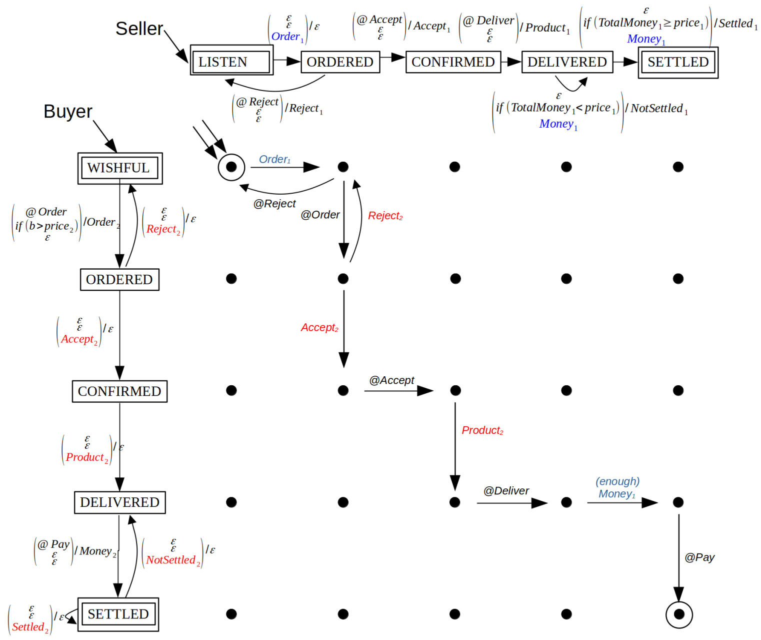

Fig. 4 A trader coordinates its two roles of buyer and seller. That is she coordinates two sequential games (trade interactions 1 and 2) through a shared budget and shared commodity counter. The decision $@Order$ in interaction 2 effectively determines the outcome of $@Accept$ in interaction 1, eliminating what appeared to be a free decision. Thus, like interaction, coordination also restricts the possible reachable state space of the product automaton of the involved roles by transition elimination and decision-dependencies.

Thus, we have shown that what seems to be a free decision in one interaction may just be the consequence of the subject being involved in some other interaction. Coordinating several roles can reduce (up to complete elimination) as well as increase the number of free decisions.

4 How we decide

The automaton model of Sections 2 – 3 specified the structure of the economic interactions. We now enrich our model with a model of our mind on how we decide, connecting the formal model to the welfare-theoretic questions this paper addresses.

4.1 A simple model of mind: preferences

If we want to hold up our model of free decisions, but nevertheless want to be able to know 'why' we act like we do, we have to come up with a model of our mind. A simple yet expressive model of our mind, which is used extensively in contemporary economics, is the preference concept. It fits very nicely in our interaction/coordination model and aligns well with standard revealed-preference interpretation where preferences are inferred from choices, not attributed independently of the decision context.

With the preference concept, we postulate that within an interaction we act as we want in the sense that we make a certain free decision because we prefer some 'outcome' over some alternative 'outcome'.

What is an 'outcome'? Intuitively, we might say that the outcome related to a decision is the target state value of the transition in which the decision is involved. But, to stay in line with game theory, it is more appropriate to relate the outcome to the interaction as a whole and identify the outcome with the success criterion of the interaction.

That is, for finite interactions, we have a set of final state values to characterize success and we make a certain free decision because we prefer one of these final state values.

Given two different outcomes a and b, we can then say that being in state $p \in Q$ and receiving some input character $i \in I^\epsilon$, with two possible transitions $p \xrightarrow{(d_1, i)/o_1} q_1$, finally leading to outcome a and $p \xrightarrow{(d_2, i)/o_2} q_2$ finally leading to outcome b, we prefer outcome a over b, notated as $a \succ_{p,i} b$ if we would take the decision d_1 that leads to a instead of d_2 leading to b. If we are indifferent to the outcome a and b we could choose with equal probability and notate $a \sim_{p,i} b$. The subscript can be dropped if it is clear from its context.

As preferences relate to outcomes of interactions, they have a natural context dependency, namely to the context of the interaction. That is, preferences in our sense need not relate, for example, apples to pears in any absolute sense, but only with respect to some interaction they are involved in.

The essential idea of the money mechanism is to extend our preference relation over those commodities that are exchangeable to a certain degree, to pairs of such commodities and *'money'*. We call these preferences *'material preferences'*.

Mathematically speaking, with M the set of money values and C the set of indexed commodities, the preference relation $\succsim \subseteq (C \times M)^2$ relates pairs of numbers of commodities and amounts of money. To keep our notation simple, we refer only to a single class of commodities.

For example, assume that we prefer to have €10 and no cinema ticket over having only €5 and such a ticket. In symbols: $(0_{\text{cinema ticket}}, €10) \succ (1_{\text{cinema ticket}}, €5)$. In this case, we would not pay more than €5 for the ticket.

Why should we as individuals do such a seemingly artificial extension of our preference relation? Because, even without reflection, we grasp intuitively the usefulness of this concept in supporting us in our challenge to coordinate our many social interactions. Part of the money we gain in one interaction we can put on the table in another interaction. We can now start to realize a lot more actions, according to our original preference relation, because of this money mechanism – why should we not prefer something that makes this possible?

So, we can state: some real-world entity that can be exchanged and can be represented by some state gets a pivotal role in our preference relation with regard to all commodities that we view as exchangeable to a certain degree. And to be able to talk about it easily, we name it *'money'*. Money in this sense works, because of its capability to support us in coordinating our diverse economically relevant interactions, and finally because we prefer it to do so as whole subjects.

4.2 Utility

As is explained in virtually every textbook on economics (for example, Mas-Colell et al. (1995)), the key idea of the utility concept is to simplify our model of mind even further by introducing a single-valued function *util*, mapping all outcomes to real numbers so that we can replace our preference relation by a comparison of these 'utility' values.

We say that a function *util*, which maps outcomes to real numbers, represents our preference relation $\succsim$, if for every two outcomes a and b, $util(a) \geq util(b)$ is equivalent to $a \succsim b$. We thereby reduce a property of a binary relation between two outcomes to a property, the utility, that can be attributed to a single outcome.

There are a number of propositions that describe under which conditions it is possible to represent preference relations with utility functions. Above all, the preference relation necessarily has to be transitive and complete. Mathematicians call such a relation a *'total order'*. In economics, such a preference relation is often termed *'rational'*. We prefer 'consistent' to 'rational' for such a preference relation, since rationality in ordinary language implies much more than transitivity and completeness.

Then, in the countable case with a fixed preference relation, there always exists a representing utility function. The same holds for continuous preference relations over convex domains (Debreu 1954). However, no such representation exists in the important case of hierarchical preference relations with infinite items on each level.

It is clear that a utility function over-specifies its corresponding preference relation. If one such utility function f exists, then any concatenation $u \circ f$ with a strictly increasing function $u \colon \mathbb{R} \to \mathbb{R}$ is also a corresponding utility function to the same preference relation.

But apart from the issue of over-specification, talking about utility (where this is allowed per assumption) is in our model the same as talking about preferences. Utility maximization in our model just means realizing what we prefer most. Thus, the popular identification of utility maximization with egoism is false in our model, since the utility concept in our model does not determine the content of the preferences it is supposed to express. Following Ernst Fehr and Urs Fischbacher (Fehr and Fischbacher 2002), egoism is much better defined as having exclusively *'selfish'* preferences that are directed only towards oneself.

So, in summary, for us utility is nothing we look for in the brains of people, but it is a mathematical concept we use to simplify our considerations on how we take our free decisions over outcomes of stateful and non-deterministic interactions, based on a preference model where this is mathematically permissible.

4.3 How buyer and seller decide in a trade

We now apply the concepts of preferences and utility to explain how a buyer B and a seller S decide in a trade interaction according to our model.

4.3.1 Buyer

As described in Section 3.1 we attributed the state $(mod, n, b, val_a, cost_a)^B$ to B with the initial values of $(\textsc{wishful}, 0, b_0, val_a, \textit{'undefined'})$.

B has the free decision $@Order$ to buy or not to buy a at costs $cost_a^B$. With her initial budget b_0^B, B would pay at most $val_a^B < b_0^B$ in exchange for a.

We now assume that B's preference relation relates pairs (b, n). B is indifferent by definition towards $(b_0^B, 0) \sim (b_0^B - val_a^B, 1)$ and – assuming 1 as the unit of money – has the following preferences:

$$(b_0^B, 0) \sim (b_0^B - val^B, 1) \prec (b_0^B - val^B + 1, 1) \prec \ldots . \tag{2}$$

Thus, we can define a utility function for the decision $@Order$, where B gains in a trade interaction the difference between her valuation val_a^B and her cost $cost_a^B$ she incurs to acquire a.

$$\text{Buyer:} \qquad util_a^B = val_a^B - cost_a^B \tag{3}$$

We see immediately that for any $cost_a^B \leq val_a^B$ this function is indeed a utility function for B, as for a given valuation val_a^B the equivalence $util_a^B(cost_1) > util_a^B(cost_2) \Leftrightarrow cost_1 < cost_2 \Leftrightarrow (1, b_0^B - cost_1) \succ (1, b_0^B - cost_2)$ holds.

Economists refer to this utility also as *'consumer rent'* or *'consumer surplus'* (e.g. Marshall (1920); Mankiw (2024)).

4.3.2 Seller

Let us turn to the seller role 'S'. According to our trade interaction, S has the free decision to accept an offer for commodity a at price p, or to reject it.

Initially, S has a certain amount of money b_0^S before the production of a. For production, S had to spend a part of his money, his $cost_a^S$. With selling, he wants at least to get compensated for his expenses. Hence, he is indifferent to the following possessions $(b_0^S - cost_a^S, 1) \sim (b_0^S, 0)$ and the more money he additionally gets in the trade, the better.

$$(b_{total}^S - cost_a^S, 1) \sim (b_0^S, 0) \prec (b_0^S + 1, 0) \prec \ldots \tag{4}$$

Thus, we can define the utility of S, he gains from a trade interaction as the difference between his $revenue_a^S$ and $cost_a^S$ to produce a.

$$\text{Seller:} \qquad util_a^S = revenue_a^S - cost_a^S \tag{5}$$

Again, it is easy to verify that this is indeed a utility function.

Economists refer to this utility also as *'profit'*, *'supplier rent'* or *'producer surplus'* (e.g. Marshall (1920); Mankiw (2024)).

4.3.3 The trade

According to our assumption of freedom of choice, both parties will only agree on this trade if both take at least some utility out of it. Hence, the valuation of B and the cost of S determine the possible price range insofar as S's cost determines the lowest and B's valuation the highest possible price. Within this price range, this trade constitutes a 'win-win' situation for both.

The total utility for both after trading a at $price_a$ is just the sum of the individual utilities and does not depend on the price.

$$\text{Total utility (trade):}\ util_{a,total} \quad = \quad util_a^B + util_a^S \quad = \quad val_a^B - cost_a^S \tag{6}$$

Note that this form of total utility depends on our assumptions of the simple form of Eqs. 3 and 5 for the utility of B and S. If we had chosen a monotonically transformed version of both then the total utility of trade would look different. For example, we could have defined $util_a^B = \log(val_a^B - cost_a^B)$ and $util_a^S = \log(revenue_a^S - cost_a^S)$. Then total utility would have been $util_{a,total} = \log\left((val_a^B - price_a)(price_a - cost_a^S)\right)$ which would depend on the price.

With the assumed utility functions of B and S, both parties contribute to the total utility. B with a high valuation and S with low costs. As in our model total utility does not depend on the price, we can attribute it to the interaction as an invariant and say: a free trade creates a utility surplus of x. The price only distributes the utility between B and S in the trade, hence the notion of money as a mechanism to distribute utility.

We give an example: A book producer and a book enthusiast have an initial budget of €100 each, making a total asset of €200. The book producer now produces a book which costs him €30, resulting in a remaining budget of €70 plus the new book. The book enthusiast would give away €90 of his €100 to get this book. Both agree on a price of €50. Subsequently, the book producer has €120 and the book enthusiast has €50 and a book he values at €90, making an asset of €140. Together, after the trade, both have a total asset of €260. The asset grew by the total utility created by the trade of €20 for the book producer and €40 for the book enthusiast. So both did benefit from the trade, though not equally.

4.4 How trader decides

Now, let us turn to our trader T, which we defined as $T = comp^{internal}(B, S)$ with a coordination condition as we described in Section 3.2. Then for B, her successful interaction 2 changes n_a by $\Delta n_a = 1$ and the budget b by $\Delta b = -p_2$ as long as $p_2 < val_a^B$ otherwise both remain constant. For S, his successful interaction 1 has $\Delta n_a = -1$ and at least $\Delta b = p_1$ as long as $p_2 > cost_a^S$.

Hence, as successful trading implies successful buying and selling, n_a remains constant and her budget will increase by $\Delta b = p_1 - p_2$, her preference relation – which concerns only outcomes – actually applies only to her budget. The more budget she gets from trading, the better.

$$\text{Utility (trader):} \quad util^T = revenue - cost = p_1 - p_2 \tag{7}$$

We achieved a perfect fit of the utilities of B, S, and T only because we ad hoc assumed the simple forms of the utilities and did not apply additional strongly monotonic transforms, possibly different for each role. This will be one of the additional requirements in the next section, where we will generalize our money model. For roles that can be combined in one subject, any possible such transformation must be uniform for all roles.

What happens if a trader trades with herself, that is, what is the result of $comp^{external}(comp^{internal}(B, S))$? In this case, nothing happens because it is not possible to sell a commodity to a higher price than to buy it. This can be generalized to all closed chains of traders:

Theorem 1 *A closed chain of traders does not perform any trade and therefore creates no utility.*

The proof is trivial as in a closed chain of arbitrary length the same holds as for a single trader who trades a good exclusively with herself: It is not possible along the whole chain to achieve a positive utility with a limited budget. This is the automaton analogue of the no-money-pump condition in revealed preference theory — a closed chain of trades must leave total utility unchanged (see also Section 7.2).

Such a closed chain is a simple example where free decisions can be enforced or prevented by external circumstances whose existence cannot be deduced from the local interaction itself. It means that any working trade (chain) assumes that in the end there are at least some real independent producers and consumers involved. More generally, it means that much of contemporary economic theory depends on the assumption of true independence of its economic subjects.

5 Money as a utility transfer mechanism

We now prove the key formal result of the paper. The game-theoretic setup of the earlier Sections, specifically the requirement that buyer and seller roles can be consistently combined in a single trader, places tight constraints on the admissible utility representations. We show these constraints uniquely determine the linear form which is the basis for utility transferability (see Harsanyi (1977), pp.111, or Myerson (1991), pp. 384)

In fact, for our aim of showing that subjective preferences become comparable with the money mechanism, it suffices to have at least one utility representation with our desired properties. Crucially, the freedom inherent in the utility concept — that no particular form is mandated — allows us to choose precisely the form that makes interpersonal comparison meaningful.

5.1 Uniqueness

We now show that the utility representations introduced for B and S are the unique ones satisfying what we call 'compositional consistency' – with respect to the external coupling of trade and the internal coupling of trader.

These constraints partition into a trade and a trader part, and each restates a commitment already made in Sections 3-4 rather than introducing a new one. Within a trade, the buyer's utility depends only on what she gains over what she paid, and normalizes to zero at break-even and to her full valuation when costs are zero. This is simply a restatement of how valuation was operationalized as a primitive of B's preference relation in Eq. (2) – the money amount at which B is indifferent to trading at all; the seller's condition mirrors this for revenue.

We want total utility to be a property of the trade and therefore independent of the price – since the external coupling condition (Eq. 1) already treats price as a pure transfer between B and S with no leakage, any additive utility measure must be invariant to how that transfer is split. Finally, as a trader, any monotone re-scaling of utility must be the same for both roles: this follows from the coordination condition of Section 3.2, where B's and S's budgets collapse into the single shared variable b^T, which retains no record of which role generated it – so no well-defined utility function of b^T can depend on a role-specific transform.

The one genuinely substantive assumption in this bundle is that utility depends only on the difference between valuation/revenue and cost, not on their levels separately: this expresses a fungibility of money across price levels that is not forced by anything stated so far. We will see in Section 6 that this is closely related to the comparable valuation condition governing when monetary comparisons across individuals are valid, and Section 10.4 catalogs circumstances – non-substitutable goods, hierarchical needs – under which it fails.

Definition 1 We name the utility functions, representing the material preferences of a buyer B and a seller S in a trade, 'compositionally consistent' if they fulfill the following constraints:

- Trade-interaction constraints:

(a) The utility of B depends only on the difference between her valuation and her costs: $util_a^B = f(val_a^B - cost_a^B)$, and becomes zero if her valuation equals her costs, $f(0) = 0$, and becomes her valuation if the costs are zero, $f(val_a^B) = val_a^B$.

(b) The utility of S depends only on the difference between his revenue and his cost: $util_a^S = g(revenue_a^S - cost_a^S)$, and it becomes zero if his revenue equals his cost $g(0) = 0$, and becomes his revenue if his costs are zero $g(revenue_a^S) = revenue_a^S$.

(c) The total utility resulting from the trade as the sum of the utility of S and B is the difference between the valuation of B and the cost of S and is therefore independent of the price: $util_{a,total} = util_a^B + util_a^S = val_a^B - cost_a^S$.

- Trader constraint: Any strongly monotonic transform of the utility functions of B and S must be identical.

Our claim can now be formulated as the following theorem.

Theorem 2 (Uniqueness) *The utility functions of a seller S and a buyer B in a trade interaction according to Def. 1, namely $util_a^S$ and $util_a^B$ as given by (3) and (5) are the only compositionally consistent utility functions.*

To prove this theorem, we first have to provide the general utility functions for S and B in a trade interaction. As we have seen, $val_a^B - cost_a^B$ is a possible utility function of Buyer and $revenue_a^S - cost_a^S$ is a possible utility function of Seller. Then, the general utility functions for B and S, which depend only on the required difference, are provided by two additional strongly monotonic functions $f, g \colon \mathbb{R} \to \mathbb{R}$ such that

$$util_a^B = f(val_a^B - cost_a^B) \tag{8}$$
$$util_a^S = g(revenue_a^S - cost_a^S) \tag{9}$$

Now, the trader constraint requires both functions to be identical: $f = g$. And the external coupling condition (Eq. 1) turns the first constraint into

$$\begin{aligned} util_{a,total} &= util_a^B + util_a^S \\ &= f(val_a^B - price_a) + f(price_a - cost_a^S) \end{aligned}$$

With the next lemma, we prove that this requires f to be linear with zero intercept.

Lemma 1. *Let f be a real monotonic function such that whenever $a, b, c \in \mathbb{R}$ with $a \geq b \geq c$ then $f(a-b) + f(b-c)$ does not depend on b. Then f is a linear function.*

We give the proof of the lemma in Appendix B.

Thus, f is a linear function. To complete the proof of our theorem, we note that together with our boundary conditions $f(0) = 0$, and $f(val_a^B) = val_a^B$, f has to be the identity function.

The uniqueness result has a strong implication: there is exactly one way to represent material preferences as utility functions that is consistent with the exchange structure. It implies that any deviation results in some loss of insights into the money mechanism. For example, if we allowed different functions f and g for the seller and buyer roles, we would lose our unifying view on both the trade interaction and the trader. This is why linearity in money is not an ad hoc assumption but a structural necessity in our model.

5.2 Measurement theory

In 1934, Oskar Lange (Lange 1934a,b) proved, corroborated by Franz Alt (Alt 1936), that if utility differences are supposed to be meaningful, then only affine transformations leave their preference representations invariant.

This posed a problem to the increasingly dominant 'marginalists', as a derivative is a limit of differences. Their answer was to separate what is being computed from what is being interpreted. They would allow themselves to compute these derivatives, but they would not allow themselves to interpret them as meaningful. Allowed for interpretation were only ratios of marginal utilities, the 'marginal rates of substitution (MRS)', as here the dependencies on the derivative of any general strongly monotone function $f = f(u)$ cancel out, since for $u = u(x, y)$, the ratio $\frac{\partial u}{\partial x} / \frac{\partial u}{\partial y}$ remains invariant against a transformation with f.

But then, as Franz Alt has already pointed out, any proposition that relates magnitudes of marginal utilities, like for example the *"law of the diminishing marginal utility"*, would also remain meaningless.

It becomes interesting, if we start to use money as the second commodity. In 1971 David H. Krantz, R. Duncan Luce, Patrick Suppes and Amos Tversky (KLST) proved as their representation theorem for conjoint structures (Krantz et al. 1971, Section 4.4, p. 151) that the presence of money as a second dimension in the preference space, together with suitable independence conditions, generates an interval-scale representation for otherwise only ordinally structured commodity preferences. Our result exploits precisely this structure, and the linear utility representation is the natural consequence.

KLST thereby establish the interval-scale structure within a single individual's originally ordinally structured preference space including money. But they do not address whether this structure can be used to compare preferences across individuals which is the contribution of our next Section 6.

5.3 Scope and clarifications

We would like to make a couple of further remarks. First, even though in the presented model the money mechanism can transfer utility in the defined sense, it does not represent or measure utility in the sense that the more money we have, the more utility we get. So, the statement *"Money is transferable utility"* is false in this model. It is true that money is a mechanism of utility transfer where individual utility manifests itself on both sides of a trade only as a difference between some sort of revenue/valuation and some sort of cost.

Second, the sentence *"The marginal utility of money decreases with increasing amount of money"* is false in this model. In fact, we do not attribute utility to money, and beside that, utility is a strict linear function of any money term.

But third, true is that our valuation per item of all commodities we buy for consumption, like apples, dramatically decreases with increasing numbers, that is, the valuations for consumption commodities are not additive. Although we may be willing to pay €1 for a single apple we want to eat now or in a few days, we will not pay

€1000 for 1000 apples in a single moment, as we might even have to pay additional money to dispose of most of them afterward.

6 Intersubjective comparison of preferences

In our model, based on the role-based nature of a trade, we can compare the preferences of different trading subjects in several different ways. We can compare the valuations of competing roles and the utility of complementary roles. And finally, we can compare the total utilities generated in different trades.

6.1 Intersubjective comparison of preferences between potential buyers

Can we compare the subjective preferences of potential buyers competing for a given commodity?

To find an answer, we recap how we defined the individual preference relation: I like the apple more than the pear, exactly when? Yes, if in a decision to consume either of them, I would choose the apple over the pear. In other words, we have actually introduced the preference as an 'explanation' of a decision that must be made in an interaction context, that is, as a 'mental model' in a given context of interaction.

If we transfer this way of talking to the relationship between two individuals, it stands to reason to refer to a decision in a corresponding context of interaction, namely the context of interaction towards a third party. Something like this: There is a third party called S, who has an apple to offer and has to decide whether to give it to A or B.

Assuming that A and B have the same budget and the same access to a comparable set of alternatives, it is obvious that S should give the apple to the one who has the higher valuation, and we agree on the meaning of the phrase 'A prefers the apple more than B' if A's valuation is greater than B's under these conditions. Thus, we define

Definition 2 Let A, B be two buyers, competing for the same commodity a in an interaction with a seller S. If both have the same budget and access to the same set of alternatives, then we define the relation 'A prefers a more than B', symbolically $A \succ_a B$, if and only if $val_a^A > val_a^B$.

We name this condition under which the valuations of two buyers become comparable the *'comparable valuation condition'*.

The introduction of a third person, namely seller S, signals the 'intersubjective objectivity' of this valuation comparison.

This is a condition on the trading environment (budgets, access), not an assumption about preferences or their comparability — in contrast to John C. Harsanyi's assumption on preference form (see Section 8.1) or Amartya Sen's assumption on comparability itself (see Section 8.2). We are neither assuming comparability to get it, nor our preferences to have some special aggregation-friendly structure. But we are deriving a common, unit-identical representation for free (any two money-using traders are automatically speaking the same currency), and then giving a separate, empirical/circumstantial condition – checkable independent of anyone's preferences – under which that shared currency also tracks equal preference intensity.

It formalizes the intuition behind Alfred Marshall's observation that willingness to pay is a reliable welfare measure only among people of equal income (Marshall 1920, pp.15–16). Daniel Goldstick already speculated in (Goldstick 1971) that under equal budgets the willingness to make a *"financial sacrifice"* is an indicator of a *"person's desire"* to acquire some economic asset. It is also the structural condition implicit in Varian's (Varian 1974) envy-freeness result: under equal budgets and equal access, any allocation resulting from free trade is envy-free and efficient (see also Moulin 2019, for a recent discussion). And it is also a validity condition for Vickrey second-price auctions (Vickrey 1961) as utility-maximizing mechanisms: the auction correctly aggregates buyer valuations into a surplus-maximizing allocation only when those valuations reflect equally unconstrained preferences.

The comparable valuation condition is like the preconditions of an ideal gas. A gas will behave like an ideal gas if the respective conditions are fulfilled. The ideal gas law is not 'assumed' to hold by kinetic theory — it is derived from a model (point particles, elastic collisions, no intermolecular forces). There is no objection to relocate the assumption that the gas is ideal into the derivation – because the derivation and the applicability condition are doing different jobs: the derivation tells us what the law looks like under those idealized conditions; the applicability condition tells us when a real gas is well-described by it, and deviations (Van der Waals corrections) are themselves informative, not embarrassing.

So, the comparable valuation condition just signals whether we are still on firm ground for this kind of money semantics or not. The more the comparable valuation condition is infringed, the worse we can compare intersubjective preferences by the willingness to pay of the buyers.

6.2 Intersubjective comparison of utility between complementary roles of B and S

Adding a buyer's and a seller's utility to a total utility in a trade implies the comparability of the utility of both roles, which we justify by the possibility to validly use the same money scale for both roles in our model.

Importantly, this allows us to define a '(local) fair' price $price_a^{fair}$ from eq. (3) and (5) as the price which provides both the same utility in their trade ($util_a^B = util_a^S$):

$$price_a^{fair} = \frac{val_a^B + cost_a^S}{2} \tag{10}$$

Eq. 10 is the Nash bargaining solution (Nash et al. 1950) for this bilateral exchange game: it distributes the surplus $val_a^B - cost_a^S$ equally between buyer and seller. This provides an axiomatic grounding for local fairness through the Nash bargaining axioms (Pareto efficiency, symmetry, independence of irrelevant alternatives, and scale invariance) without additional assumptions.

It is immediately apparent that if a price can be fair, it can also be arbitrarily unfair. Although a free trade is, by definition, a win-win situation, it can nevertheless be highly inequitable, in the sense that the party with greater bargaining power may capture nearly all of the total surplus. We thereby get a glimpse of the power of money as a mechanism to distribute utility: we might be able to design it such that it distributes utility by and large fairly, or we might design it willingly or unwillingly to become a mechanism to distribute utility systematically in an extremely unfair manner.

However, fairness is a rather complex concept and, to quote Hervé Moulin (Moulin 2019), *"There is no single compelling definition [of fairness]"*, which we would like to underline by what we distinguish as 'local' versus 'global' fairness in trades. With local fairness, we refer to a single trade. With global fairness, we refer to many trades.

6.2.1 Global fairness: fairness in 1:n relations

What happens if a seller sells his commodities not to a single buyer but to n buyers? The first thing we have to realize is that, in reasoning about this case, we aggregate the utility of different subjects. The validity of our reasoning therefore depends on the validity of our assumptions, that everyone has the same total budget and has the same access to all relevant alternatives. In this 1:n case, the order of utility partition and price determination plays a decisive role for fairness.

The first possibility is to average the utility first and then to distribute it evenly by deriving the individual prices thereof. Let us again look at the book producer who now produces 2 books at €30. There are two book enthusiasts, one with a valuation of €90, the other, a bit less enthusiastic, with a valuation of €60. The average total utility then is the total valuation minus the total cost divided by the number of participants (€90 + €60 - €30 - €30)/3 = €30. If the first book is sold for €60 and the second book for €30, every participant gets off individually with the average utility of €30. Formally we have:

Step 1 [Determine average utility]: $\bar{u} = \frac{1}{n+1}\sum_i (val_i - cost_i)$
$= \frac{n}{n+1}\left(\overline{val} - \overline{cost}\right)$

Step 2 [Determine n prices]: $p_i = val_i - \bar{u}$

Thus, for increasing n, the average utility that everyone realizes converges to the difference between the average valuation and the average cost.

The second possibility is to determine the local 'fair' price in each trade directly and afterward look at the utilities of every participant. As S participates in every trade interaction, he now gets his share n times. The first price would be determined at €60 and the second price at €45. The utility of the first buyer would be €30, of the second buyer €15 and for our seller it would be €45. Formally this is:

Step 1 [Determine n prices] $p_i = (val_i - cost_i)/2$

Step 2a [average utility of the n buyers] $\bar{u}^B = \frac{1}{n}\sum_i (val_i - p_i)$
$= \overline{val} - \bar{p}$

Step 2b [utility of seller] $u^S = \sum_i (p_i - cost_i)$
$= \sum_i u_i^S = n\bar{u}^S$

Now, all buyers get on average the difference between their average valuation and the average price, and the seller's total utility increases with the number of buyers (or sold items), and thereby shows the same scaling behavior as with a constant price for all trades.

The overall utility and therefore also the average utility for all participants is identical in both cases, but the distributions are different.

Our result shows that 'average first, price second' allocation gives equal utility to all participants while the 'price first, average second' method systematically advantages the seller as it shows the same scaling behavior as with a constant price for all trades.

This distinction becomes of great relevance in a company employing n employees or in the case of a small number of organizations producing a large number of goods for very many consumers.

6.3 Aggregation of utilities of different trades: markets

With the fulfillment of the comparable valuation condition, we can directly compare all buyers' valuations. Together with the objectively determinable costs of all sellers and our requirement that the total utility of a single trade is independent of the price, this makes a market mechanism maximize the total utility of all possible trades.

To prove this, we just have to order all sellers according to their increasing cost and all buyers according to their decreasing valuation, as shown in Fig. 5. The resulting 'market price' p^* creates the largest enclosed area and thereby maximizes total utility. This diagram can already be found in the influential textbook of Alfred Marshall 'Principles of Economics' (Marshall 1920, p. 668).

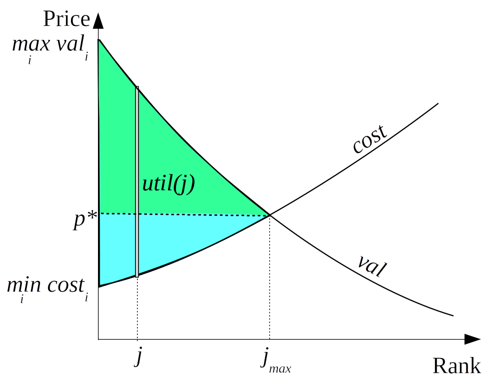

Fig. 5 Supply and demand of the market mechanism. All sellers are indexed according to their increasing cost and all buyers according to their decreasing valuation. The intersection of both curves determines the 'market price' p^* together with the index j_{max}.

The comparable valuation condition therefore provides a validity condition for the demand side of the market. On the supply side an obvious and well-known validity constraint is that the costs of the suppliers are the true costs for society and no externalities exist. We call a market mechanism that maximizes utility in the sense of all, a *'valid'* market mechanism. Only valid markets fulfill Friedrich A. Hayek's (Hayek 1945) conception *"to coordinate the separate actions of different people in the same way as subjective values help the individual to coordinate the parts of his plan."*

Whether this optimum also maximizes welfare depends on whether people do prefer what benefits them most (see Section 10.2). This as well as the validity property is obviously a property of the market context and not of the market mechanism in isolation.

Be aware that a market could create different prices for all trades and would still maximize total utility, as long as the prices converge to p^* for the rank moving towards j_{max}.

With any validity condition being violated, we can still technically add all differences between the individual valuations and costs to some value, and the market mechanism will still generate some price that maximizes this value, but this price will not relate in any simple sense to the preferences of the participants. So we are not allowed to speak about this value as a 'maximized total utility'.

In our opinion, the violation of the market validity condition is one of the most common reasons for (mostly unnoticed) severe market failure nowadays, where the powerful market mechanism enforces the interests of a very few at the expense of the many, under the guise of 'freedom of choice'.

A prominent example is monopsony in labor markets, where the employer faces no comparable budget constraint relative to workers, making monetary comparison of their valuations invalid. A second is markets for non-substitutable goods (life-saving medicines, housing in constrained cities), where the hierarchy of needs violates the substitutability assumption underlying the preference model. A third is the dynamic ticket prices for the FIFA Club World Cup 2026, where a functioning market mechanism fails to allocate the tickets to those who like watching football matches most, but is designed to maximize revenue of FIFA.

6.4 Invariance and the informational basis of comparability

As the allowed utility transformations can be used to classify different levels of expressiveness (see Section 8.2), we would like to briefly mention that our money-metric utility is a ratio-scale full comparability (RFC), that is, the following proposition holds:

Proposition 3 *Let* $util_a^{B_i} = val_a^{B_i} - cost_a^{B_i}$ *and* $util_a^{S_j} = revenue_a^{S_j} - cost_a^{S_j}$ *be the utility representation of Theorem 2 of the* i*-th buyer and* j*-th seller. The only transformations of this profile that preserve its derivation are common positive rescalings* ${util_a^{B_i}}' = \alpha\, util_a^{B_i}$ *and* ${util_a^{S_j}}' = \alpha\, util_a^{S_j}$ *with a common factor* $\alpha > 0$*, identical for all* i *and* j*.*

Proof We know from our proof of Theorem 2 that the functions f_i and g_j from Eq. (8) and (9) are the identity functions for each individual, which are homogeneous to degree one and therefore the transformation ${util_a^{B_i}}' = \alpha_i\, util_a^{B_i} = \alpha_i f_i(val_a^{B_i}, cost_a^{B_i}) = f_i(\alpha_i\, val_a^{B_i}, \alpha_i\, cost_a^{B_i})$ holds; similar for $util_a^{S_j}$ for a factor β_j.

To keep the comparable valuation condition invariant, any transformation must keep equal buyers' budgets equal, therefore all α_i must be equal.

Now it remains to be shown that also all β_j equal this α – which is the case as the trade condition requires $cost_a^B = revenue_a^S = price_a$ for all trades. $\square$

7 Quasi-linear utility

The quasi-linear model is standard in mechanism design and social choice. It underpins the Vickrey-Clarke-Groves mechanism, the Myerson-Satterthwaite impossibility result, and large parts of matching theory. Our derivation shows that this model is not merely a mathematical convenience but follows necessarily from the structure of monetary exchange.

7.1 Relation to the traditional 'quasi-linear' utility model

The traditional approach in economics towards utility (for example Mas-Colell et al. 1995, pp. 310)) focuses solely on the preference relation of the consumer, assuming a utility function $u = u(x)$ of a single consumer[8], mapping to a so-called 'consumption bundle' $x = (x_1, \ldots, x_L)$ where L is the number of goods and each $x_l \in \mathbb{R}_0^+$.

The budget b of our consumer lets it access more or less goods, that is, it defines the volume of its budget set space by the constraint $b \geq p \cdot x$ for any price vector $p \in \mathbb{R}^L$.

Léon Walras (Walras 1874) introduced the simplifying idea to partition the consumption set into a single good of interest x_1 and to lump the consumption rest $(x_2, \ldots, x_L)$ into a 'numeraire' component y with an ad hoc assumed proportional utility with price 1.

We thus get the utility representation for the consumer $u = u(x_1, y) = y + val(x_1)$ with $x_1, y \in \mathbb{R}_0^+$, which economists traditionally term 'quasi-linear' to denote the linear, additive separable contribution of the numeraire to the utility.

This numeraire y is not 'money' per se, but as a composite good priced 1 per unit it is still subject to the budget constraint that $b \geq p_1 \cdot x_1 + y$. Assuming exhaustion of the budget, i.e. equality, we can thus eliminate y in our utility representation by substituting $y = b - p_1 \cdot x_1$. That is, we can interpret the numeraire y as the amount of money left over after acquiring the bundle x_1 at price p_1. The utility then becomes

$$u = u(x_1, b, p_1) = b + val(x_1) - p_1 \cdot x_1. \tag{11}$$

This is, apart from a different offset, the utility form of Eq. 3 for our buyer. For decisions the offset is irrelevant as it is only the utility difference between the state 'before' and 'after' the trade that guides the decision of our buyer.

As Eq. 11 implies the resulting utility-maximizing demand x_1^* to be reached at $p_1 = \frac{\partial\, val}{\partial x_1}(x_1^*)$. As this expression is independent of b, it is usually said that the quasilinear utility representation does not show any wealth effects. That is, the optimal choice of x_1 is supposed to be determined by the condition that the marginal utility of

[8] we do not mention the index j for the consumer to simplify notation

x_1 equals its marginal cost in terms of money foregone. If you give consumers an extra €100, they spend it entirely on the numeraire leaving their demand for x_1 unchanged.

However, in our model, the valuation (or willingness to pay) is the behaviorally primitive object and there is no reason to assume it not to depend on the available budget as $val = val(x_1, b)$ and therefore $p_1 = \frac{\partial\, val}{\partial x_1}(x_1^*, b)$.

Despite this dependence, our model of trade still qualifies as a Myerson's game of transferable utility where (Myerson 1991, p. 384) *"each player i has the option to give any amount of money to any other player, or even to simply destroy money, and each unit of net monetary outflow or loss decreases i's utility payoff by one unit"* as the money flow between buyer and seller is part of the essential core of this interaction.

So, for a single trade, $val(x_1, b)$ contributes nothing beyond what $val(x_1, b^*)$ for some fixed budget b^* already gives us — it is just the classical model with a constant plugged in. But looking beyond a single trade, we are in fact asking about a family of games, indexed by b. And across that family the worth of the coalition $\{B, S\}$, fails to be a single, parameter-free number: v depends on which member of the family you are in, because $\partial val \partial b \neq 0$.

So, it seems to us that the classical model is not precise enough with respect to its context, indicating the local character of the utility concept.

In summary, we grounded the standard quasi-linear treatment of the consumer utility (as a buyer) with our explicit interaction-theoretic model not as an ad-hoc assumption but as a structural consequence of the composition constraints, but only locally for single (trade) interactions.

7.2 The mechanism design perspective: Revealed Preferences and empirical evidence for linear utility in the money term

According to our model, our preferences are mirrored in our decisions. In what he described as an attempt to drop *"the last vestiges of utility analysis"* Paul A. Samuelson (Samuelson 1938) introduced the concept that was later named 'revealed preferences' (see e.g. Chambers and Echenique 2016, for an overview) by postulating as a consistency condition of a buyer's choice: *"if an individual selects batch one over batch two, he does not at the same time select two over one."*

Hendrik S. Houthakker (Houthakker 1950) proved in 1950 that the possibility to base the choices of a buyer on a utility function is equivalent to the fulfillment of the 'Strong Axiom of Revealed Preference (SARP)'[9]. It states that if a buyer directly or indirectly prefers bundle x to bundle y, she can never, under any circumstances, prefer bundle y to bundle x. That is, there cannot be any cycles in utility-representable preferences.

In 1967, Sydney Afriat (Afriat 1967) was the first to provide a procedure to construct a utility function directly from finitely many customer choices x_i at prices p_i, $i = 1, \ldots, N$ with linear programming techniques, the 'Afriat inequalities'. If this data set was free of revealed preference cycles containing a strict relation, then there exist numerical values ('utility levels') u_i and positive weights λ_i, such that $u_i - u_j \leq \lambda_j(p_j(x_i - x_j))$ for all i, j.

[9]This term was attributed by Paul A. Samuelson in (Samuelson 1950) in response to Hendrik S. Houthakker's publication.

But why should people tell us their preferences? Why should not they lie or – more likely – state something different because of strategic reasons? The answer is provided by game theory, and more specifically by the discipline of mechanism design, as it was developed by Leonid Hurwicz, Eric Maskin, and Roger Myerson and others. Mechanism design can be understood as the part of game theory which is concerned with the optimal choices of the rules of social interactions or 'games' - or 'mechanisms'.

Mechanism design provides the important 'revelation principle' which states that if we know a certain mechanism to exist, then there also exists a respective direct mechanism where the participants find it always optimal to truthfully report their decision-relevant private information or 'type' as it is often termed in the literature.

Building upon this work, Jean-Charles Rochet showed in (Rochet 1987) that under quasilinearity, a mechanism is truthfully implementable in dominant strategies if and only if the decision rule satisfies cyclical monotonicity. The proof essentially uses the property of the assumed utility being transferable via money in a linear way. Thereby monetary transfers become a clean instrument to align incentives. Intuitively it says that no money pump exists.

Donald J. Brown and Caterina Calsamiglia (Brown and Calsamiglia 2007) used these results in a consumer choice setting to develop a test for quasilinear preference. Given a data set of pairs of prices and chosen quantities (p_t, x_t), $t = 1, \ldots, N$, these choices can be explained by a continuous, concave and strictly monotone quasilinear utility function if and only if the data set is 'cyclically monotone'.

Marco Castillo and Mikhail Freer (Castillo and Freer 2023) showed that these assumptions are equivalent to cyclical monotonicity on the linearized budget sets and extended these results to nonlinear budgets, which they assumed to be downward closed[10] and describable by a continuous and decreasing[11] function μ_t, such that $b_t = \{(x, m) \in Y : \mu_t(x) \geq m\}$.

Economically these assumptions mean that (i) if you can afford (x, m) by choosing x, you can certainly afford to have less money left over. And consuming less of everything while retaining the same money is also within the budget set.

So cyclical monotonicity plays the same formal role in both contexts — it is the condition that prevents any 'profitable cycling' through alternatives, whether those alternatives are other type reports (Rochet 1987) or other feasible bundles (Castillo and Freer 2023)

However, the empirical part of the article of Marco Castillo and Mikhail Freer (Castillo and Freer 2023) is problematic in two respects. First, it relates to a (simple) work task, and as we have seen, our attitude towards our work might be governed also by some 'internal valuation'. And secondly it involved only 65 undergraduate students.

We thereby contribute to the theory of revealed preference, as we say that with the money mechanism we can gather even interpersonal information about the relation of material preferences of different people. However, systematic empirical testing of quasilinear preferences in exchange settings remains an open research agenda.

[10] A downward closed budget set is a budget set with the property: if $y \in B$, then for every $y' \leq y$ also $y' \in B$, where '$y < y'$' means for the vectors $y_i \leq y'_i$ for all componentes and $y_i \neq y'_i$, that for at least one component.

[11] A function $\mu : X \to \mathbb{R}$ is decreasing if $x \geq x'$ implies $\mu(x) \leq \mu(x')$ and $x > x'$ implies $\mu(x) < \mu(x')$).

8 Other approaches to economic interpersonal comparability

The standard position in mainstream economics is that any interpersonal comparison of preference requires extra-normative structure. In the following we present the approaches of John C. Harsanyi's aggregation assumptions and Amartya Sen's social welfare functionals.

8.1 Harsanyi's expected-utility aggregation theorem

John C. Harsanyi (Harsanyi 1955), extending the work of Marcus Fleming (Fleming 1952), assumed a structural equivalence between individual and social (aggregated) preferences under risk, namely that both fulfill the expected utility axioms and can be represented as von Neumann-Morgenstern utility functions U_j and U, respectively (von Neumann and Morgenstern 1944).

Then, according to John C. Harsanyi's assumption, the preferences of each individual $j = 1, \ldots, n$ can be expressed as $U_j(p) = \sum_{i=0}^{m} p_i U_j(e_i)$ and the social preference as $U(p) = \sum_{i=0}^{m} p_i U(e_i)$, both summing over all outcomes.

He additionally required that the aggregated preferences should be indifferent between a pair of lotteries when all individuals are indifferent between them, what he called 'Pareto indifference'.

John C. Harsanyi's social aggregation theorem[12] then says that the aggregation of the 'social' preference relation can be viewed as a linear superposition of all individual (utility) preference relations. Formally, this means that there exist $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $\mu \in \mathbb{R}$ such that $U(p) = \sum_{j=1}^{n} \lambda_j U_j(p) + \mu$, now summing over all individuals.

To guarantee non-negative coefficients λ_j, the Pareto indifference condition has to be augmented to entail not just indifference, but the 'larger than' relation, called 'Semistrong Pareto condition': for all $p, q \in \mathcal{L}$, if $U_j(p) \geq U_j(q)$ for all $j = 1, \ldots, n$, then $U(p) \geq U(q)$.

To guarantee a unique solution (λ, μ), the individual utility functions U_j have to be affinely independent, which means that the solution (λ, μ) to the equation $\sum_{j=1}^{n} \lambda_j U_j(p) + \mu = 0$ is only $\lambda_1 = \cdots = \lambda_n = \mu = 0$.

Actually John C. Harsanyi's original proof was based on the implicit assumption of affine independence. But as other authors have shown, this assumption is not necessary (for a recent overview, see Chakravarty et al. 2022, ch. 7).

However, the theorem says neither how the individual utilities are to be determined nor how to determine the social weights. For the latter, John C. Harsanyi introduces ad hoc a second, 'ethical' preference relation per individual, namely that of an 'impartial observer' who faces the additional lottery over which identity she will assume in the set of individuals, which results in an equal probability distribution.

Unlike Harsanyi's approach, which derives comparability from assumed structural equivalence of individual and social preferences under risk, the present paper derives it from the structure of bilateral exchange. No assumptions about preferences under risk or lottery equivalences are required.

[12] This term was introduced by John A Weymark Weymark (1991). John C. Harsanyi named it 'Theorem V' in his 1955 publication

8.2 Amartya Sen's social welfare functionals

In 1951 Kenneth Arrow published his now-famous dissertation 'Social choice and individual values' (Arrow 1951) where he proved the impossibility to aggregate individual preferences to a 'social preference' in a way to fulfill certain sensible properties, assuming the impossibility to interindividually compare these preferences beyond their simple ranking (p.59).

In his critical engagement with these ideas, Amartya Sen (Sen 1970, 2017) developed his concept of the 'social welfare functionals' (Ch. 8*2, p. 185).

Instead of assuming, like Kenneth Arrow, that individual ranking preferences are to be aggregated by a 'social welfare function' (his terminology), Amartya Sen proposed to use the 'utility'-functions, mapping 'social states' onto real values as starting point and to derive the ranking relation thereof by a 'social welfare functional', allowing for a richer theoretical framework of measurability and comparability assumptions.

Tweaking the requirements for the equivalence conditions can be used to distinguish ordinal and cardinal measurement conditions as well as different intersubjective comparability conditions (see d'Aspremont and Gevers 2002, for a recent summary). On this basis, Amartya Sen developed the concept of partial comparability (Sen 1970).

Focusing on partial comparability, Amartya Sen reframes the problem of interpersonal aggregation of welfare indicators as being inherently incomplete, value-laden, and context-dependent. He replaces the overly pessimistic 'impossibility' result with a more realistic view of what social evaluation can be achieved – while accepting that some indeterminacy is unavoidable.

Later Claude d'Aspremont and Louis Gevers extended Amartya Sen's work in d'Aspremont and Gevers (1977) further and defined classes of admissible utility transformations, asking which social welfare orderings remain invariant under those transformations. Kevin W. S. Roberts (Roberts 1980) then provided a rigorous representation theory for informational assumptions, characterize all compatible social choice rules. Together their contribution was to make Amartya Sen's concept of social welfare functionals mathematically exact, exhaustive, and usable as a general framework for axiomatic welfare economics.

To summarize, in Amartya Sen's framework, interpersonal comparability is an input, a parameter of the social welfare functional that the analyst specifies. In the present paper it is an output, a derived property of monetary interaction under the comparable valuation condition. As shown formally with proposition 3, our conceptualization of money in fact supports the ratio-scale full comparability (RFC) condition, of which cardinal full comparability is a special case in the d'Aspremont-Gevers framework; the two aggregation results of Sections 6.2 – 6.3 are based on this property.

9 The relation to traditional welfare economics

In his highly influential textbook 'The Economics of Welfare', Arthur C. Pigou draws a strong connection between our welfare and the purpose of economics as a science: (Pigou 1920, p. vii) *"The complicated analyses which economists endeavour to carry through are not mere gymnastic. They are instruments for the bettering of human life. The misery and squalor that surround us, the dying fire of hope in many millions of*

European homes, the injurious luxury of some wealthy families, the terrible uncertainty overshadowing many families of the poor— these are evils too plain to be ignored. By the knowledge that our science seeks it is possible that they may be restrained. Out of the darkness light! To search for it is the task, to find it, perhaps, the prize, which the "dismal science of Political Economy" offers to those who face its discipline."

Welfare economics suffers particularly from any restriction of intersubjective comparison. It is simply not possible to do welfare economics without weighing the welfare of different subjects against each other (Drakopoulos 2024).

At the beginning of the 20th century, leading economists like Alfred Marshall, who pioneered the use of consumer surplus as a monetary measure of consumer welfare, still believed that utility could be cardinally measured, as did Arthur C. Pigou who treated utility as a quantifiable variable and assumed that interpersonal comparisons of well-being were valid and necessary.

With the rise of scientific positivism and behaviorism, this perspective came, as we mentioned already, to be increasingly discredited within mainstream economics as 'unscientific' and the view spread that the ordinal structure of our preferences prohibits any intersubjective comparison.

9.1 Pareto efficiency: avoiding interpersonal utility comparison

Vilfredo F. Pareto (1848–1923) was the first to make the distinction between *'ordinal'* and *'cardinal'* utility and introduced the idea to handle the analysis of economic equilibria with ordinal utility (Aspers 2001).

Following his contributions, a feasible allocation of resources is now termed *'Pareto-efficient'* if no other such allocation exists that makes some consumer better off without making some other consumer worse off (Mas-Colell et al. 1995, p. 313). In this sense, a Pareto-efficient state is a state that cannot be changed further on a strictly voluntary basis, as it provides every single participant with a veto against unfavorable changes from its perspective. Obviously, an intersubjective comparison of preferences is unnecessary to determine Pareto efficiency.

So, Pareto efficiency says very little about the desirability of any distribution. Even allocating all wealth to a single person could result in a Pareto-efficient outcome after the fact, if this single person prefers not to share.

The supposed relevance of the concept of Pareto efficiency was underscored by the designation of two so-called 'welfare theorems' until the 1950s which were based directly on this concept. The first states that under ideal conditions (complete information, perfect competition) market mechanisms lead to a Pareto-efficient outcome, that is a price vector that equalizes supply and demand (e.g. Mas-Colell et al. 1995, Section 16C). The second states that any Pareto-efficient allocation can be reached as a competitive equilibrium for some initial set of endowments (e.g. Mas-Colell et al. 1995, Section 16D). Together they indicate that the concept of Pareto efficiency is equity-agnostic in that equity and Pareto-efficiency are not necessarily trade-offs.

But, as both theorems are actually only faintly related to 'welfare' in a colloquial sense, we share the view of Martin J. Osborne and Ariel Rubinstein (Osborne and Rubinstein 2023, p.131ff) who refrain from using these terms, as they *"regard the word 'welfare' as inappropriate in the context of the result."*

For Pareto efficiency to tell us anything sensible about distribution, it would be essential that the initial allocation against which improvements are measured has already been reached on a completely voluntary basis. However, as history shows, this is rarely, if ever, the case. In fact, it is well-known from both a theoretical and empirical perspective that Pareto efficiency is a poor model for human decision-making (Myerson 1991, p. 98).

To give some simple examples: Why should we not take away the money that a thief has stolen? Why should we not redistribute money distributed by mechanisms realizing any fairness model only to a very limiting extent, perhaps even invalidating our desired money semantics to make our material preferences intersubjectively comparable? Why should we not be able to correct the effects of erroneous preferences, even against the beneficiaries of the errors?

The insight that the Pareto-efficiency property of market mechanisms – that is, the way they distribute the social payout – depends on the assumption of total voluntariness also implies that the design and enforcement of markets cannot be delegated to the market mechanism itself.

We agree with Amartya Sen (Sen 2008) that the wide acceptance of the Pareto criterion in economics as a model for *'rational behavior'* seemed to be a direct consequence of the denial of the possibility to interpersonally compare preferences: economics became a doctrine to justify the economic status quo, one that obscured the potentially severe societal consequences of invalid markets and even removed the issue of social justice from the economic agenda, culminating in Friedrich Hayek's dismissal of social justice as *"strictly empty and meaningless"* (Hayek 1976, p. 68).

Interestingly, if we accept Hayek's conclusion – that the impossibility of intersubjective preference comparisons entails the non-existence of social justice – then by contraposition, the very meaningfulness of social justice implies that such comparisons must be possible. Accepting the obvious relevance of social justice as a concept, we can – with the help of Friedrich Hayek's valid conclusion – *prove* that interpersonal comparison of subjective preference must be somehow possible. Our result provides one such form, grounded in exchange structure rather than normative stipulation.

9.2 Valuation as 'willingness to pay (WTP)'

In standard welfare economics, WTP is used to measure how much money an individual would give up for a good or change and thereby infer preference intensity from monetary choices. It underlies tools like cost–benefit analysis or compensating / equivalent variation.

But traditionally, WTP is viewed as problematic for interpersonal comparison, as €10 from a poor person is not equal to €10 from a rich person in welfare terms.

As WTP does not measure 'utility' directly but also reflects budget asymmetries, the standard critique is therefore, that it cannot serve as a common scale across individuals. Hence, willingness to pay is viewed as a pragmatic but theoretically problematic approach.

With our model, we see that the problem with WTP is not fundamental, but it arises from unequal economic conditions. We show that if you impose the condition of 'comparable valuation', then WTP becomes a valid signal for material preference intensity across individuals.

This vindicates Alfred Marshall who – 100 years earlier – emphasized, *"the money which people of equal incomes will give to obtain a benefit or avoid an injury is a good measure of the benefit or injury"* (Marshall 1920, p.15–16). And it supports Eli Cook (Cook 2025) who coined the term *"Willingness to Pay Without Apology"* for his critique of this ignorance of wealth effects, citing N. Gregory Mankiw's highly influential contemporary textbook as a prototype (Mankiw 2024, p. 134).

As a consequence, this perspective places the limits of the valuation concept into the center of the discussion, as our money model inherits all strengths and limitations of the preference model. To make money do its work, we do not have to assume stable, ever lasting, globally complete and transitive preference. We only need sufficiently structured and representable preferences within any given interaction context.

The money mechanism only claims to operationalize whatever preference structure exists. So: if preferences are context-sensitive, money becomes context-sensitive; if preferences are socially constructed, money operationalizes socially constructed valuation; and so forth. The monetary layer cannot be 'cleaner' than the preference layer beneath it.

As pointed out by Cass R. Sunstein (Sunstein 2025), our valuation can vary enormously in time for a variety of reasons. Amartya Sen underlined in (Sen 2017, p. 30) the *"malleability of mental attitudes"*. Kenneth Galbraith emphasized the creation of wants in his book 'The Affluent Society' (Galbraith 1958) where he calls (p.156) *"modern want creation"* to be *"the most obtrusive of all economic phenomena"*. And for preference construction and context-dependent preferences, see also (Lichtenstein and Slovic 2006; Tversky and Simonson 1993).

We would also like to mention Marc Fleurbaey and François Maniquet's (Fleurbaey and Maniquet 2011) theory of fair monetary valuations that is consistent with ordinal preferences and budget constraints. Our result is complementary: we show that monetary valuations support meaningful interpersonal comparisons specifically when budget constraints are equalized — the same condition that fair income equivalent approaches implicitly require.

One widely used analytical tool within welfare economics that is based on WTP is cost-benefit analysis (CBA) which evaluates policies by summing up individuals' WTP and subtracting their collective costs. As CBA rests on the same implicit assumptions as WTP it inherits the same vindication under the comparable valuation condition.

For the important difference between collectively maximizing our preference fulfillment and the much more complex aspect of collectively maximizing welfare, we refer to our discussion in Section 10.2.

9.3 Measuring utility differences with money metric functions

John R. Hicks (Hicks 1939) introduced so-called 'money metric functions' to measure utility differences with money, either by the amount of money needed to compensate a subject to reach its original utility after a change in prices, termed 'compensating variation', or by the amount of money needed to achieve an equivalent utility change as the actual price change, evaluated at old prices, termed 'equivalent variation'.

Despite the fact that money-metric utility provides mappings $m_i(x)$ only for the i-th individual to their personal income equivalent of utility in state x, these values are usually added together to calculate a social welfare function $W(x) = \sum_i m_i(x)$.

Our contribution is to deliver the economic conditions, namely the 'comparable valuation condition' which justifies this summing by making money a common yardstick across individuals, eliminating the relevant economic differences between people.

10 Starting points for broadening the perspective of welfare economics

The preceding sections established the formal core of our argument: money, understood as a role-based interaction mechanism, generates intersubjectively comparable valuations under precisely characterizable conditions. Accepting this result, however, forces a set of broader questions that standard welfare economics has largely set aside.

Section 10.1 situates our money model within a philosophy-of-science perspective on what models are and do, clarifying both the force and the limits of our uniqueness result. Section 10.2 addresses the most consequential implication: since preferences model free decisions rather than well-being, market-aggregated utility measures comparative want-satisfaction, not welfare in any richer sense. Section 10.3 extends the trade model to the employment relationship, where the seller acquires a discretionary "internal valuation," yielding new insights into reservation wages and recursive fairness preferences. Section 10.4 examines substitutability as the foundational assumption of the one-dimensional utility representation: where it fails, the money mechanism loses its comparability properties. Finally, Section 10.5 explores the tension between money's "tool" function as a social coordinator and its "drug" function as a reward mechanism capable of tainting the preferences it is meant to express — framed through the complementary principles of Athena-Invariance and Hermes-Desire.

10.1 Models as our mean to differentiate between true and false

To fully understand the relevance of our contribution, a short excursion on our understanding of a model is probably helpful.

We understand a model in the sense of mathematical model theory (for example, Ebbinghaus et al. 1996) as what is needed to evaluate a proposition as being true or false. Thereby 'having a model' becomes a necessary precondition to 'understand something'. Models in this sense do not 'mirror' a somehow given reality, as is often formulated (see, for example, Argyris 1960, p. 5), but represent what of the world is accessible to our mind (German: 'Verstand') in this sense. Models themselves are not true or false, but only more or less appropriate or practically useful. Apart from the requirement not to permit the derivation of contradictory propositions, we thereby have a certain leeway in how we understand the world.

Thus, we cannot force anybody to accept our money model, but if we do not adopt it as the base for our understanding of how the money mechanism operates and opt for some other model, we have to live with the consequences, as we lose certain – in our view quite important – explanatory features.

Under this perspective, the fact, that there is just one way to represent our preferences as a utility function that fulfills all the requirements of our interaction/-coordination model of the economic subject does mean that there is exactly one way to understand money if we want to use it to intersubjectively compare preferences in the proposed way.

10.2 Where to place 'utility': free decisions versus well-being

Another far-reaching consequence of this model perspective is that we have to decide which meaning we want to attribute to the term 'utility'. Economics has long used the word 'utility' in at least two senses. The first, following Samuelson's revealed-preference program, takes preferences and thereby utility as the explanatory component of what a subject chooses, agnostic about whether the choice is good for the subject or not. The second, going back to Bentham and revived in the 'experienced utility' literature (Kahneman et al. 1997), takes utility to represent what is actually good for the subject — her well-being, independent of what she happens to choose.

Daniel Hausman argues in (Hausman 2011) that these two notions are routinely conflated in welfare economics, and that this conflation is not harmless. Amartya Sen made the choice-theoretic half of the same point earlier in 'Rational Fools' (Sen 1977), arguing that a purely choice-based concept of preference is too thin to ground welfare judgments.

We introduced preferences to fill the leeway provided by the nondeterminism of our interactions as a model for how we decide, very much in the sense of Sarah Lichtenstein and Paul Slovic (Lichtenstein and Slovic 2006), who even think that *"preference construction is now viewed as the core problem in decision making"*.

They are therefore *not* a measure of our well-being. In our model, the meaningful statement *"I prefer A over B because A is more in my true self-interest than B"* is not a tautology. In fact, any identification of our preferences with our well-being would be a contradiction to our assumption of freedom of decision, as it is an essential element of this assumption that we have the ability to take decisions that are obviously – at least from the point of view of almost any external observer – against our genuine self-interest. That is, in our model, the proposition 'I prefer A over B, because A is more in my true self-interest than B' is *no* tautology.

It follows that price signals, willingness to pay, and the total utility maximized by a valid market are claims about aggregated comparative wants, but not about welfare. And what people comparatively want depends greatly on their social context, experience, and conditioning. It seems possible to bring quite intelligent people, put in some appropriate social context, to agree almost unequivocally on the strangest purposes. As an extreme example: When Joseph Goebbels publicly asked, at a mass rally in Berlin's Sportpalast on February 18, 1943 – only weeks after the German military catastrophe of Stalingrad – "Do you want total war?", he received an unequivocal, massive 'Yes!'.

One likely reason is that the preference model grasps the structure of our mind only very superficially. Following Daniel Kahneman (Kahneman 2011) and distinguishing a fast, instinctive and a slow, thoughtful modus of thinking, a simple way of manipulating our preferences would be to address our instinctive over our thoughtful way of thinking. Important examples for such manipulations can be found in the advertising industry or in the so-called 'social media'. Especially the latter even amplifies our instinctive thinking on a global scale by many feedback mechanisms, like presenting more content that has shown to be effective in triggering instinctive thinking.

Statements like *"consumer surplus measures well-being gains"* are not just empirically contestable, as critics have long noted – within our model they are category errors, since they ask money to do something only the welfare interpretation of utility could license, while money's comparability properties (Theorem 2) depend on the decision interpretation holding. The composition-consistency result that makes money a meaningful comparison device is the same result that forbids reading its output as direct measure for welfare.

In one sense, our result delivers what welfare economics has long sought: a structural grounding for interpersonal preference comparisons. In another, it forecloses the very interpretation that would make such comparisons directly welfare-related – since the same model that licenses comparison prohibits reading its output as a direct welfare measure.

We actually see this as a strong justification for welfare economics to broaden its scope in the sense of Arthur C. Pigou which is already ongoing, as the extensive discussion demonstrates about indicators other than GDP to better assess societal prosperity in a more holistic sense (e.g. Stiglitz et al. 2009).

In our opinion, the impossibility to interpret preference-based utility as a direct measure for welfare and have the proposed money model at the same time, signals that welfare economics is not about determining the one and only real-valued parameter which has to be optimized and everything is fine forever. But instead it should also focus on these validity questions: Are our market mechanisms really valid? Do our preference really provide evidence concerning what makes us better off? What is the role of economics in a broader social picture of hierarchically preferred things like freedom, education, etc.?

Not taking into account that aggregated preference are not necessarily a welfare indicator incurs a substantial risk that *"cost-benefit analysis and other techniques of welfare economics may be misapplied"*, as Daniel Hausman (Hausman 2011, p. 136) formulated.

10.3 Extension to trading labor

One important topic, a broadened welfare economics should definitely focus on, is the relevance of our labor for our welfare. This leads us to a simple but important extension of our money-based trade model if we look at the simple question: how do we trade our labor?

As Karl Marx suggested (Marx 1849, p. 399), like sugar? That is, as any other substitutable commodity without reference to our genuine interest in the task? Then, our current model of a trade would suffice: as a seller of our labor, that is, as an employee, our salary would be our revenue, our expenses for reproduction would be our costs, the difference would be our utility, and we would be completely indifferent towards what we do. The less we work the larger would be our utility. As a logical conclusion, collective utility would be maximized, if we could eliminate labor altogether – the authors would never have taken the effort to write this article.

Does this make sense? How do we delineate sensibly between reproductive and other costs? How do we classify child-related expenses? Why does this article exist? Compared to selling an apple, which takes place in a virtual second, our labor time is an essential part of our life. Should our model not be able to accommodate our preferences for certain types of work over others?

10.3.1 The discretionary component of the employee's decision

One approach to find answers to these valid questions is to simply extend our model with a discretionary component also on the seller's side, the employee, symmetric to that of the buyer, the employer.

Then, the employee's cost term no longer reflects purely objective facts; instead, it represents an 'internal valuation' — a minimum acceptable salary below which the employee would decline the work. The more the employee prefers the job, the lower the minimum salary the employer will at least have to pay for it.

This minimum acceptable amount of money is also called the 'reservation wage' in the literature (Mortensen 1977). A low reservation wage signals a strong preference for the job, provided that the decision is truly free. Under forced circumstances, it may just imply the fact that the person desperately needs money 'at all costs'.

This simple model has some interesting consequences. First, labor markets do not work the same way as markets for ordinary goods, which is already common sense in behavioral economics (Dohmen 2014). The internal valuation also plays a significant role in the explanation of unemployment (Krueger and Mueller 2016).

It also explains the fact that the higher an employee (internally) values his job, the larger some incentive, like severance pay, must be for him to give it up voluntarily.

Further, as with selling ordinary goods, the total utility can be influenced by both participants. The employer by creating a productive environment where it makes economically sense to pay an employee a lot of money – that sounds ok; and the employee by preferring this work so much that he is willing to do the job for almost nothing, even negative amounts are conceivable – that sounds strange. At first glance, it may seem that this inverse relation enforces an indemnification semantics on every salary where high salaries can only be achieved if they compensate for all the scourges of the job.

But in reality, bad jobs are usually poorly paid. And on a second glance we see that the valuation of the employee does represent only the lower boundary of the possible salary range and not the salary itself. The actual salary will depend to a large extent on the price determination mechanism – and in particular on two recursive mechanism, that we discuss in the next section.

10.3.2 Two recursive mechanisms determining our preference for work

We would like to mention two possible effects of recursive preference relations[13], both of which possibly have important effects on salary levels.

We first consider an employee with a higher-order preference for local fairness. Let us assume that an employer has to offer a job that she values at €130,000. The employee has an initial valuation of €10,000. Knowing little about each other, the employer and the employee agree on an initial salary of €20,000 and the employee starts working.

After a while, the employee discovers that the employer's valuation is in fact €130,000 and, therefore, a fair salary would have (initially) amounted to €70,000. What sum would this employee have to demand in order to remain in the company? If he claims €70,000 as the minimum salary to stay in the company, he, in fact, relates to his own, older valuation of €10,000 – which has just changed to €70,000. So, assuming €70,000 as minimum acceptable salary, a fair salary would amount to €100,000, and so on and so forth. The preference relation becomes recursive and the fixed point is the employer's valuation, in this case €130,000.

Formally, the following proposition holds:

Proposition 4 *Let Er be an employer and Ee an employee. Er values the work of Ee with* val^{Er}*. Ee values his own work initially with* $val_0^{Ee} < val^{Er}$*. Additionally, Ee has a hierarchical fairness preference, such that he does not work for a salary that he perceives as unfair according to eq. (10).*

Assuming Ee knows val^{Er} *then a recursive process starts where the negotiated salary has* val^{Er} *as fixed point, resulting formally in* $lim_{n\to\infty} util_n^{Ee} = lim_{n\to\infty} util_n^{Er} = 0$*.*

Proof For the proof we calculate the employee's valuation from eq. (10) in the $n-th$ step as $val_n^{Ee} = val^{Er} \sum_{i=1}^{n-1} \frac{1}{2^i} + val_0^{Ee} \frac{1}{2^{n-1}}$. As the sum of the first summand converges to 1, the fraction of the second summand converges to 0, the stated proposition holds for the limit $n \to \infty$.

□

In the limit, the internal valuation of the employee corresponds exactly to the external valuation of the company. The utility of both employer and employee shrinks to zero. This just means that both the employee and the employer achieve what they prefer most: for the employee that he is treated fairly and for the employer that she reaches her business goals – the usual case for organizations that are not owned by someone, like associations, foundations, etc.

This interesting constellation shows nicely the limit of the concept of a one-dimensional utility in case of hierarchical preferences, in this case with respect to our

[13] According to our knowledge, recursive preferences were introduced to economics by David M. Kreps and Evan L. Porteus (Kreps and Porteus 1978) and Larry Selden (Selden 1978) to distinguish the separate roles of time and risk. See also the work of Larry G. Epstein and Stanley E. Zin (Epstein and Zin 1989) and Philippe Weil (Weil 1990).

labor. How should welfare economics take this into account, that a utility value of zero signals an optimal outcome?

We think that this leads the way to a completely different interpretation of the employee's remuneration. We can understand employment not as a periodic trade of labor versus money, but as a membership in an organization which the employee becomes a logical part of, and the remuneration not as costs for the organization, but as the employee's share of the organization's collectively generated utility – an interpretation that is strongly supported by the role of the employees' salaries in the GDP-calculation, and which could explain the huge horizontal wage dispersion across different organizations to a certain extent (Mortensen 2023).

Another recursive relation, this time between the preferences of the employer and the employee, might be created by a dependency of the external valuation of the employer on the internal valuation of the employee. The employer might expect that an employee who highly values doing a task will on average perform this task better than another one with lower internal valuation. In this case, it would make sense for the employer to valuate employees with higher internal valuation higher than employees with lower internal valuation, although the latter would possibly start working only for higher salaries.

10.4 Substitutability

As we showed with our price determination for our work under local fairness conditions, utility representations of our preferences break down under hierarchical conditions. In fact, John S. Chipman pointed out in 1960 (Chipman 1960) that substitutability is the essential property for a one-dimensional real-valued utility measure. So, another essential structural aspect of our money model with its one-dimensional real-valued utility measure is therefore the substitutability of all goods where our monetarily expressible preferences are directed at.

Indeed, the illegitimately enforced interchangeability of things that are actually hierarchically preferred permits the extraction of arbitrarily high prices, capped only by the buyer's budget: what would we think of a skipper who, in the middle of the ocean, spots a man adrift, clinging to a plank, and only begins rescuing him after the wretched 'voluntarily' hands over his entire fortune? According to our money model, invoking the concept of utility in such a case to justify prices is simply a blatant misuse of economic terminology to disguise outright plunder.

This example is not that far-fetched. In Germany, the pharmaceutical industry can unilaterally set the price of a new drug for the first year. For new, so-called 'personalized therapies', the industry justifies prices of up to one million euros not with the costs, but with the 'benefit' it creates for the individual patient (Röhrlich 2018).

In fact, with the money mechanism, we partition our world into what can be coordinated through exchange and what resists coordination through exchange. With the money mechanism we not just compare values, but we define what counts as comparable. Money constructs a partial space of comparability, and outside it there is no coherent scalar measure of trade-offs at all. Thus, the scope of comparability is itself an outcome of institutional structure in the sense of Karl Polanyi (Polanyi 1944) who emphasized that commodification is always socially constructed and thereby politically contestable.

As a concrete policy implication regarding economic inequality: the greater the inequalities, the more cautious we should be in using money as a welfare metric or market-based allocation if fairness is important.

And finally, there are many, highly valued things, many of them 'immaterial', that we cannot buy directly, like health, education, freedom, democracy, peace, etc. Thus, from a monetary perspective, we live in an 'indirect society'.

10.5 The drug versus tool effect of money

As most economists would probably agree, without truly free decisions, the whole field of economics would become superfluous. By contraposition, the evident vitality of our economy could be interpreted as evidence of truly free decisions. It is no coincidence that especially economists seem to stress our capability to make truly free decisions. Actually, the dominance of the Pareto-efficiency concept in economics could be interpreted as a grotesquely exaggerated relevance of our freedom of decision.

As freedom of choice is something people value very highly, and economic theory actually presupposes it, to preserve it should be a natural target for welfare economics. What circumstances enhance, and what circumstances diminish our decision freedom are obviously very relevant economic questions.

The conceptual impossibility to fill the leeway of our freedom of decision with our self-interest indicates the complexity of this topic. Additionally, as we have seen: what seems a free decision in one interaction could be, and in fact often is, determined by within-subject coordination with another interaction.

When some of our economic interactions start to determine our actions, the first thing we lose is our social coordination ability. Poverty is a good example, as Karl Marx accurately noted when he wrote in (Marx 1894, p. 828): *"The realm of freedom begins, in fact, only when the work determined by necessity and external expediency comes to an end*[14]*."* In the same direction goes Amartya Sen's understanding of poverty as capability deprivation (Sen 2017, p. 25).

The very fact that money can both strongly support social coordination and, through the same mechanism, drag us into desolate slavery[15] is itself strong evidence for the relevance of the subject of freedom of choice for economics.

Taking a step back, we find another link between money and the subject of our freedom of choice noticing that this freedom can not only be severely abridged by external coercions, but also, without doubts, internally, namely by drug addiction as described by Wolfram Schultz (Schultz 2011). In our model of the mind, the disastrous consequences of drugs are reflected in a manipulative change in the preferences of the drug addicted, we name 'tainting'.

What has money to do with drugs? Stephen E. G. Lea and Paul Webley (Lea and Webley 2006) distinguished between the 'drug aspect' and the 'tool aspect' of money. The tool aspect is clear, as it comprises its support of our coordinative abilities.

[14] German original: *"Das Reich der Freiheit beginnt in der Tat erst da, wo das Arbeiten, das durch Not und äußere Zweckmäßigkeit bestimmt ist, aufhört."*

[15] A more academic expression would perhaps be 'engender severe dependency', but this does not convey the appropriate existential connotation to remove virtual all individual freedom.

But these authors argue that there remain some phenomena related to our behavior towards money that can much better be accounted for by assuming that money works by mimicking *"the neural, behavioural, or psychological action of some other, more natural incentive"*.

How does money affect our preferences according to our model? There are two well-known simple consequences from the linear utility representation of money (see e.g. Mas-Colell et al. 1995, p. 45) that relate directly to the tension of the money concept as described by Stephen E. G. Lea and Paul Webley. To emphasize their importance, we name one with 'Athena', goddess of wisdom in ancient Greece, who unwaveringly pursues what is right and is naturally not swayed by money and gold – and the other one with 'Hermes', god of trade and merchants.

Athena-Invariance: if someone is indifferent with respect to two bundles $x, y \in \mathbb{R}^n$, $x \sim y$, then any additional amount of money (component 1) does not change that:

$$(x + \alpha e_1) \sim (y + \alpha e_1), \forall \alpha \in \mathbb{R}, e_1 = (1, 0, ..., 0) \tag{12}$$

Hermes-Desire: Money (component 1) is desirable in the sense that everything else constant, we prefer more money over less money:

$$(x + \alpha e_1) \succ (x), \forall \alpha > 0 \tag{13}$$

If money worked according to our model, then by Athena-Invariance it should not influence our preference relation of material goods and by Hermes-Desire, ceteris paribus it should always be better to have a bit more than a bit less money. These two principles do not contradict, but complement each other well in a delicate balance and capture, what our money-intuition is about.

One known mechanism by which money can taint our preferences is to work as a reward. In psychology, a reward is defined as an external stimulus that is supposed to reinforce an otherwise spontaneously occurring behavior in simple learning situations, called classical (Pavlovian) and operant conditioning[16]. Newer research shows, that it is the discrepancy ('prediction error') between the reinforcer that is predicted based on a stimulus and the actual reinforcer which determines learning by reward (Waelti et al. 2001; O'Doherty et al. 2003).

A reward's effectiveness for learning is bound to a certain degree of freedom necessitating the internal construction of the prediction model. It entails a strong tendency to shift the motivation to act from internal sources to the reward as an external source (Deci and Ryan 2000), which is, by definition, a change in our preferences.

Scientists have identified the neural subsystem concerned with this type of learning and call it the *'reward system'* (Schultz 2015) and here is where the respective drugs unfold their catastrophic effect. It is a core part of our complex autoregulation system providing us with our autonomy, why we find the term *'autoregulation system'* much more appropriate.

[16]To contrast this phenomenon to an externally provided confirmation of the veracity of a subject's behavior and thereby a reinforcement of the subject's internal source of motivation, the latter could be named a 'recognition'.

Due to Hermes-Desire, money is particularly well-suited as a reward. And due to Athena-Invariance we perhaps have a sensitive schema to detect preference tainting. From this perspective, the broad introduction of reward-oriented compensation schemes in the tradition of the highly influential economists Michael C. Jensen and William H. Meckling (Jensen and Meckling 1976, 1994), especially for top management – holders of the most consequential positions in the private economy, becomes highly problematic, a fact that is also intensively discussed in the literature (e.g. Frey and Osterloh 2005).

As truly free decisions play such an important role for the functioning of our economy, it would be worthwhile for welfare economics as a science to come up with systematic approaches for fostering this capacity in us and certainly not to diminish it.

Statements and Declarations

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgement We would like to thank Marvin Deversi and Leoni Bossemeyer who gave very valuable feedback to earlier versions of this paper. Also Martin Kocher, Christiane Schwieren for their encouragement to put these ideas on paper and Stavros Drakopoulos for his helpful hints on the literature. And finally, it was Christine Reich, who had the idea to assign the term 'valuation' also to the cost-term of the employee with respect to its work and thereby emphasizing the symmetric relation between the external valuation of the work by the employer and its inner valuation by the employee.

Appendix A Document classes, modes, and parameters for transitions

To describe more complex roles efficiently, we introduce a certain abstraction by summarizing the elements of the transition relation in equivalence classes (Reich 2022). While this may appear somewhat technical, the practical example in Section 3 should render it straightforward. Its main idea is that we can describe the interactions not by the exchange of characters, but by the exchange of documents, that we are all familiar with.

The key idea is to create equivalence classes that preserve the time structure of the transitions by identifying a combination of document classes, the most relevant state value components, and additional conditions in a way that we can still attribute one set of entities to time t (like i and p) and the other entities to $t' = successor(t)$ (like o and q).

We introduce three injective functions:

1. $parse_{IO} \colon I \cup O \to DocCls \times Param_{IO}$ assigns each character $a \in I \cup O$ a document class $docCls(a) \in DocCls$ and a value of a parameter set $param_{IO}(a) \in Param_{IO}$. We can thereby regard each I/O character of a decision system as an element of some document class that represents certain parameter values.
2. $parse_Q \colon Q \to Q_{mode} \times Q_{rest}$ assigns each state value $q \in Q$ a mode component $mode(q) \in Q_{mode}$ and a rest component $rest(q) \in Q_{rest}$.
3. $parse_D \colon D \to DCls \times Param_D$ assigns each decision value $d \in D$ a decision class component $dCls(d) \in DCls$ and a value of a parameter set $param_D(d) \in Param_D$. We can thus regard each decision as an element of a given decision class that represents certain parameter values.

And we introduce a set of conditions $cond$ as functions $cond_{dCls(d),mode(p),docCls(i)} : Param_D \times Q_{rest} \times Param_{IO} \to \{true, false\}$, which, for each decision class, mode and document class, evaluates the values of the remaining state and the parameters of the decisions and input documents to either true or false.

With these tools, we now define the desired equivalence relation:

Definition 3 Two transitions $t, t' \in \Delta$ are equivalent, symbolically $t \sim_\Delta t'$, if their values for the decision and document classes, the modes, and the condition are equal, that is, iff $dCls(d) = dCls(d')$, $docCls(i) = docCls(i')$, $docCls(o) = docCls(o')$, $mode(p) = mode(p')$, $mode(q) = mode(q')$, and
$cond_{dCls(d),mode(p),docCls(i)} = cond_{dCls(d'),mode(p'),docCls(i')}$.

Thus, with $dCls(d)$, $mode(p)$, $docCls(i)$, and $cond_{dCls(d),mode(p),docCls(i)}$ we explicitly relate the transition equivalence to all possible values of (p, d, i) at time t and with $mode(q)$ and $docCls(o)$ we ignore $rest(q)$ and $param(o)$ at all times $t' = succ(t)$.

Instead of the set of all transitions of a single equivalence class Δ_l we also write:

$$mode(p) \xrightarrow{(dCls(d), docCls(i), cond_{dCls(d),mode(p),docCls(i)})/docCls(o)} mode(q) \tag{A1}$$

To help recognize decisions, mode values, and document classes, we prefix all decision names with an '@', put all mode values in small caps, and write all document classes with an initial capital.

We can now say that, for example, a seller who receives a purchase order document *Order* from a customer decides *@Accept* to accept the order and passes accordingly from mode value LISTENING to ORDERED, provided that the customer is assessed as *trustworthy*, and, accordingly, returns a confirmation document *Confirmation* and notate this as:

$$\text{LISTENING} \xrightarrow{(\mathit{@Accept}, \mathit{Order}, \mathit{isTrustworthy})\,/\,\mathit{Confirmation}} \text{ORDERED}$$

It is important to keep in mind that, unlike the $p \stackrel{i/o}{\rightarrow} q$ notation, which denotes a single transition, Eq. (A1) denotes a whole class of equivalent transitions.

Appendix B Proof of Lemma 1

Here, we supplement the proof of our main technical lemma which we skipped in the main text body.

Lemma 1. *Let f be a real monotonic function such that whenever $a, b, c \in \mathbb{R}$ with $a \geq b \geq c$ then $f(a-b) + f(b-c)$ does not depend on b. Then f is a linear function.*

Proof If we choose $b = (a+c)/2$ we see that

$$f(a-b) + f(b-c) = 2f((a-c)/2)\ . \tag{B2}$$

This value even depends only on $a-c$ and not on a and c individually. As $f(a-b)+f(b-c)$ is supposed not to depend on b at all, equation (B2) holds for every b.

Substituting $x = a-b$ and $y = b-c$ we can rewrite this as

$$\frac{f(x)+f(y)}{2} = f\left(\frac{x+y}{2}\right) \qquad \text{for all } x, y \geq 0\ . \tag{B3}$$

That is, the mean of the function values of x and y equals the function value of the mean of x and y. This is certainly fulfilled by every linear function, but conversely it is also true that every monotonic function fulfilling this condition is linear.

We first show the linearity for rational coefficients and in a second step extend this result to general real numbers. W.l.o.g. let $f(0) = 0$ (The general case with $f(0) \neq 0$ can be reduced to this one by taking $g(x) = f(x) - f(0)$ and looking at g instead of f.)

$f(nz) = nf(z)$ for all $z \geq 0$ and $n = 0, 1, 2, \ldots$ follows by induction: The cases $n = 0$ and $n = 1$ are trivial, by our assumption $f(0) = 0$. For the inductive step, let $n \geq 1$ and set $x = 2nz$ and $y = 2z$ in equation (B3). Then the right-hand side simply is $f\left(\frac{2nz+2z}{2}\right) = f((n+1)z)$, whereas the left-hand side becomes $\frac{f(2nz)+f(2z)}{2} = \frac{2nf(z)+2f(z)}{2} = nf(z) + f(z) = (n+1)f(z)$ (using the inductive hypothesis).

By substituting z with $\frac{z}{n}$ in $f(nz) = nf(z)$, we get $f\left(\frac{z}{n}\right) = \frac{f(z)}{n}$ for all $z \geq 0$ and $n = 1, 2, \ldots$. Combining this with the above we obtain $f(qz) = qf(z)$ for any rational coefficient $q = \frac{m}{n}$ (with $m, n = 1, 2, 3, \ldots$). In particular, we have $f(q) = qf(1)$ for any rational q.

To extend the result to general, real coefficients r we use the monotonicity of f: Namely we have $f(r) \geq f(q) = qf(1)$ for any rational $q \leq r$ which shows that $f(r)$ cannot be smaller than $rf(1)$ and, likewise, we find that $f(r)$ cannot be larger than $rf(1)$ either, hence $f(r) = rf(1)$. □

References

Afriat SN (1967) The construction of utility functions from expenditure data. International Economic Review 8(1):67–77. https://doi.org/10.2307/2525382

Alt F (1936) Über die Messbarkeit des Nutzens. Zeitschrift für Nationalökonomie/Journal of Economics pp 161–169. Translated by Siegfried Schach as "On the Measurability of Utility" and printed 1971 in J.B. Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein (Eds.) *"Preferences, Utility, and Demand - a Minnesota Symposium"* p. 424–431, Harcourt Brace Jovanovich, Inc.

Argyris C (1960) Understanding organizational behavior. The Dorsey Press, Inc. Homewood, Illinois

Arrow KJ (1951) Social choice and individual values. New Haven, New York / London: J. Wiley / Chapman & Hall, reprinted as: Arrow, Kenneth J. (1963). Social choice and individual values (2nd ed.). New Haven: Yale University Press.

Aspers P (2001) Crossing the boundary of economics and sociology: The case of Vilfredo Pareto. American Journal of Economics and Sociology 60(2):519–545. https://doi.org/10.1111/1536-7150.00073

Borgatti SP, Foster PC (2003) The network paradigm in organizational research: A review and typology. Journal of Management 29(6):991–1013. https://doi.org/10.1016/S0149-2063(03)00087-4

Börgers T (2015) An introduction to the theory of mechanism design. Oxford University Press

Brown DJ, Calsamiglia C (2007) The nonparametric approach to applied welfare analysis. Economic Theory 31(1):183–188. https://doi.org/10.1007/s00199-006-0087-5

Castillo M, Freer M (2023) A general revealed preference test for quasilinear preferences: theory and experiments. Experimental Economics 26(3):673–696

Chakravarty SR, Mitra M, Mutuswami S (2022) Social Aggregations and Distributional Ethics. Cambridge University Press

Chambers CP, Echenique F (2016) Revealed Preference Theory. 56, Cambridge University Press

Chipman JS (1960) The foundations of utility. Econometrica: Journal of the Econometric Society pp 193–224. https://doi.org/10.2307/1907717

Cook E (2025) Money talks: The rise of willingness to pay without apology. History of Political Economy 57(S1):141–174. https://doi.org/10.1215/00182702-12134446

d'Aspremont C, Gevers L (1977) Equity and the informational basis of collective choice. The Review of Economic Studies 44(2):199–209. https://doi.org/10.2307/2297061

d'Aspremont C, Gevers L (2002) Social welfare functionals and interpersonal comparability. Handbook of Social Choice and Welfare 1:459–541

Debreu G (1954) Representation of a preference ordering by a numerical function. Decision processes 3:159–165

Deci EL, Ryan RM (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55:68–78

Desai N, Mallya AU, Chopra AK, et al (2005) Interaction protocols as design abstractions for business processes. IEEE Trans Software Eng 31(12):1015–1027. https://doi.org/10.1109/TSE.2005.140

Dohmen T (2014) Behavioral labor economics: Advances and future directions. Labour Economics 30:71–85. https://doi.org/10.1016/j.labeco.2014.06.008

Drakopoulos SA (1989) The Historical Perspective of the Problem of Interpersonal Comparisons of Utility. Journal of Economic Studies 16(4). https://doi.org/https://doi.org/10.1108/EUM0000000000136

Drakopoulos SA (2024) Value judgements, positivism and utility comparisons in economics. Journal of Business Ethics 189(3):423–437. https://doi.org/10.1007/s10551-023-05395-z

Ebbinghaus HD, Flum J, Thomas W (1996) Einführung in die mathematische Logik, 4th edn. Spektrum Akademischer Verlag GmbH

Epstein LG, Zin SE (1989) Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica 57(4):937–969. https://doi.org/10.1142/9789814417358_0012

Farwer B (2001) Omega-Automata. In: Grädel E, Thomas W, Wilke T (eds) Automata, Logics, and Infinite Games, Lecture Notes in Computer Science, vol 2500. Springer, pp 3–20, https://doi.org/10.1007/3-540-36387-4_1

Fehr E, Fischbacher U (2002) Why social preferences matter - the effect of non-selfish motives on competition, cooperation and incentives. The Economic Journal 112:C1–33. https://doi.org/10.1111/1468-0297.00027

Fleming M (1952) A cardinal concept of welfare. The Quarterly Journal of Economics 66(3):366–384. https://doi.org/10.2307/1885309

Fleurbaey M, Maniquet F (2011) A theory of fairness and social welfare, vol 48. Cambridge University Press

Frey BS, Osterloh M (2005) Yes, managers should be paid like bureaucrats. Journal of Management Inquiry 14(1):96–111. https://doi.org/10.1177/1056492604273757

Galbraith JK (1958) The affluent society. Houghton Mifflin Harcourt

Goffman E (1959) The presentation of self in everyday life (dt.: Wir alle spielen Theater). University of Edinburgh Social Sciences Research Centre

Goldstick D (1971) Assessing utilities. Mind 80(320):531–541

Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63(4):309–321. https://doi.org/10.1086/257678

Harsanyi JC (1977) Rational behavior and bargaining equilibrium in games and social situations. Cambridge University Press

Hausman DM (2011) Preference, Value, Choice, and Welfare. Cambridge University Press

Hayek FA (1945) The use of knowledge in society. The Americal Economic Review 35(4):519 – 530

Hayek FA (1976) Law, Legislation and Liberty, Vol. 2, The Mirage of Social Justice. University of Chicago Press

Hicks JR (1939) Value and Capital, 2nd edn. Oxford at the Clarendon Press

Hicks JR (1989) A Market Theory of Money. Oxford University Press, https://doi.org/10.1093/0198287240.001.0001

Holzmann GJ (1991) Design and validation of computer protocols. Prentice-Hall, Inc., Upper Saddle River, NJ, USA

Houthakker HS (1950) Revealed preference and the utility function. Economica 17(66):159–174

Jensen MC, Meckling WH (1976) Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics 3:305–360

Jensen MC, Meckling WH (1994) The nature of man. The Journal of Applied Corporate Finance pp 4–19

Jevons WS (1957) The Theory of Political Economy, 5th edn. Macmillan & Co, the first edition was published in 1871.

Kahneman D (2011) Thinking, fast and slow. Allen Lane, London

Kahneman D, Wakker PP, Sarin R (1997) Back to Bentham? Explorations of experienced utility. The quarterly journal of economics 112(2):375–406

Krantz DH, Luce RD, Suppes P, et al (1971) Foundations of Measurement Volume I. Academic Press, republished 2006, Dover Publications, isbn:978-0-486-45314-9

Kreps DM, Porteus EL (1978) Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46(1):185–200. URL http://www.jstor.org/stable/1913656

Krueger AB, Mueller AI (2016) A contribution to the empirics of reservation wages. American Economic Journal: Economic Policy 8(1):142–179. https://doi.org/10.1257/pol.20140211

Lange OR (1934a) The determinateness of the utility function. The Review of Economic Studies 1(3):218–225. https://doi.org/10.2307/2967485

Lange OR (1934b) Notes on the Determinateness of the Utility Function: III. The Review of Economic Studies 2(1):75–77. https://doi.org/10.2307/2967554

Lea SE, Webley P (2006) Money as tool, money as drug: The biological psychology of a strong incentive. Behavioral and Brain Sciences 29(2):161–176. https://doi.org/10.1017/S0140525X06009046

Lichtenstein S, Slovic P (2006) The Construction of Preference: An Overview. In: Lichtenstein S, Slovic P (eds) The Construction of Preferences. Cambridge University Press, p 1 – 40

Mankiw NG (2024) Principles of Economics, 10th edn. Cengage

Marshall A (1920) Principles of Economics, 8th edn. Palgrave Macmillan, reprinted 2013 by Palgrave Macmillan in its 'Classics in Economics' series

Marx K (1849) Lohnarbeit und Kapital. Marx Engels Werke Band 6, S. 397-423, Dietz Verlag Berlin, 1959

Marx K (1894) Das Kapital, Band III. Marx Engels Werke Band 25, Dietz Verlag Berlin, 1964

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic Theory. Oxford University Press

Mortensen DT (1977) Unemployment insurance and job search decisions. ILR Review 30(4):505–517. https://doi.org/10.1177/001979397703000410

Mortensen DT (2023) Wage Dispersion: Why are Similar Workers Paid Differently. MIT Press

Moulin H (2019) Fair Division in the Internet Age. Annual Review of Economics 11:407–441. https://doi.org/10.1146/annurev-economics-080218-025559

Myerson RB (1985) Bayesian Equilibrium and Incentive Compatibility: An Introduction. In: Hurwicz L, Sonnenschein DSH (eds) Social Goals and Social Organization: Essays in Memory of Elisha Pazner. Cambridge University Press, p 229–259

Myerson RB (1991) Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, Massachusetts, reprinted 1997 as paperback

Nash JF, et al (1950) The bargaining problem. Econometrica 18(2):155–162

von Neumann J, Morgenstern O (1944) Theory of Games and Economic Behavior. Princeton University Press

O'Doherty JP, Dayan P, Friston K, et al (2003) Temporal difference models and reward-related learning in the human brain. Neuron 28:329–337

Osborne MJ, Rubinstein A (2023) Models in Microeconomic Theory, 2nd edn. Open Book Publishers, Camebridge, UK, https://doi.org/10.11647/OBP.0362

Pigou AC (1920) The Economics of Welfare. MacMillan, London

Polanyi K (1944) The great transformation: the political and economic origins of our time. Beacon Press, Boston, originally published: New York: Farrar & Rinehart, 1944 and reprinted in 1957, 2001 by Beacon in Boston.

Reich J (2022) Verwirrende Informatik I – Systeme, Informationen, Berechenbarkeit und Daten. Informatik Spektrum 55(6):372–380. https://doi.org/https://doi.org/10.1007/s00287-022-01502-y

Reich J (2023) Verwirrende Informatik II – Interaktion und Koordination von Systemen. Informatik Spektrum 46:79–86. https://doi.org/https://doi.org/10.1007/s00287-022-01519-3

Robbins L (1938) Interpersonal comparisons of utility: a comment. The Economic Journal 48(192):635–41. https://doi.org/10.2307/2225051

Roberts KWS (1980) Interpersonal comparability and social choice theory. The Review of Economic Studies 47(2):421–439. https://doi.org/10.2307/2297002, URL https://doi.org/10.2307/2297002, https://academic.oup.com/restud/article-pdf/47/2/421/4514777/47-2-421.pdf

Rochet JC (1987) A necessary and sufficient condition for rationalizability in a quasi-linear context. Journal of mathematical Economics 16(2):191–200. https://doi.org/10.1016/0304-4068(87)90007-3

Röhrlich D (2018) Neue Medikamente: Hoffnungsträger und Preistreiber. Deutschlandfunk

Samuelson PA (1938) 'a note on the pure theory of consumer's behaviour' and 'addendum'. Economica 5(19):61–71, 353–354. Reprint in 'The Foundations of Price Theory Vol 4', pp. 101–116, 2024, Routledge

Samuelson PA (1950) The problem of integrability in utility theory. Economica 17(68):355–385

Schultz W (2011) Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. Neuron 69(4):603–617. https://doi.org/10.1016/j.neuron.2011.02.014

Schultz W (2015) Neuronal Reward and Decision Signals: From Theories to Data. Physiological reviews 95(3):853—-951. https://doi.org/10.1152/physrev.00023.2014

Selden L (1978) A new representation of preferences over "certain x uncertain" consumption pairs: The "ordinal certainty equivalent" hypothesis. Econometrica: Journal of the Econometric Society pp 1045–1060. https://doi.org/10.2307/1911435

Sen A (1970) Interpersonal aggregation and partial comparability. Econometrica: Journal of the Econometric Society pp 393–409. https://doi.org/10.2307/1909546, reprinted as Ch. 9 in A. Sen (1982) Choice, Welfare and Measurement, Basil Blackwell - Oxford

Sen A (1999) Development as Freedom. New York, Alfred A. Knopf

Sen A (2008) The Discipline of Economics. Economica 75:617–628. https://doi.org/10.1111/j.1468-0335.2007.00660.x

Sen A (2017) Collective Choice and Social Welfare, Expanded Edition. Penguin Random House UK, reprint as expanded version of the original version from 1970

Sen AK (1977) Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. Philosophy & Public Affairs pp 317–344

Shannon CE (1948) A Mathematical Theory of Information. Bell System Technical Journal 27:379–423, 623–656

Stiglitz JE, Sen A, Fitoussi JP (2009) Report by the Commission on the Measurement of Economic Performance and Social Progress

Sunstein CR (2025) Willingness to pay. Foundations and Trends in Microeconomics 13(1):1–67. https://doi.org/10.1561/0700000073

Tversky A, Simonson I (1993) Context-dependent preferences. Management science 39(10):1179–1189

Varian HR (1974) Equity, envy, and efficiency. Journal of Economic Theory 9(1):63–91. https://doi.org/10.1016/0022-0531(74)90075-1

Vickrey W (1961) Counterspeculation, Auctions, and Competitive Sealed Tenders. The Journal of Finance 16(1):8–37. https://doi.org/10.2307/2977633

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412(5):43–48

Walras L (1874) Éléments d'économie politique pure, ou, Théorie de la richesse sociale. Corbaz

Weil P (1990) Nonexpected utility in macroeconomics. The Quarterly Journal of Economics 105(1):29–42. https://doi.org/10.2307/2937817

Weymark JA (1991) A reconsideration of the Harsanyi–Sen debate on utilitarianism. In: Elster J, Roemer JE (eds) Interpersonal comparisons of well-being. Cambridge University Press, p 255 – 320