Justice as a Social Bargain and Optimization Problem

Andreas Siemoneit

Carl von Ossietzky Universität Oldenburg, Germany.
ZOE Institute for future-fit economies, Bonn, Germany.

June 2021

Abstract: The question of “Justice” still divides social research and moral philosophy. Several Theories of Justice and conceptual approaches compete here, and distributive justice remains a major societal controversy. From an evolutionary point of view, fair and just exchange can be nothing but “equivalent”, and this makes “strict” reciprocity (merit, equity) the foundational principle of justice, both theoretically and empirically. But besides being just, justice must be effective, efficient, and communicable. Moral reasoning is a communicative strategy for resolving conflict, enhancing status, and maintaining cooperation, thereby making justice rather a social bargain and an optimization problem. Social psychology (intuitions, rules of thumb, self-bindings) can inform us when and why the two auxiliary principles equality and need are more likely to succeed than merit would. Nevertheless, both equality and need are governed by reciprocal considerations, and self-bindings help to interpret altruism as “very generalized reciprocity”. The Meritocratic Principle can be implemented, and its controversy avoided, by concentrating on “non-merit”, i.e., institutionally draining the wellsprings of undeserved incomes (economic rents). Avoiding or taxing away economic rents is an effective implementation of justice in liberal democracies. This would enable market economies to bring economic achievement and income much more in line, thus becoming more just.

Keywords: Justice, reciprocity, meritocratic principle, self-binding, economic rents.

License: Creative-Commons CC-BY-NC-ND 4.0.

Highlights:

• Reciprocity is primary principle of justice, only complemented by need and equality
• Justice must be just, but also effective, efficient, and communicable
• Need and equality are governed by reciprocal considerations
• Self-bindings help to interpret altruism as “very generalized reciprocity”
• Avoiding or taxing away economic rents is an effective implementation of justice


Contents

1 Introduction 3

2 Reciprocity as the Foundational Principle of Justice 4
   2.1 The Sociobiology of Cooperation 4
   2.2 The Meritocratic Principle 6

3 Justice as a Social Bargain and Optimization Problem 8

4 Extending the Picture 9
   4.1 Self-Binding 9
   4.2 Moral Intuitions 10
   4.3 Efficient Decision-Making 11

5 Need as Auxiliary Principle 11
   5.1 The Domain of Need 11
   5.2 Need in the Family? 11
   5.3 The Role of Deservingness for Welfare 12
   5.4 Summary 13

6 Equality as Auxiliary Principle 13
   6.1 The Domain of Equality 13
   6.2 Equality as Default 13
   6.3 Equality in Groups 13
   6.4 Equality of Access 15
   6.5 Political Equality 15
   6.6 Summary 15

7 Practices of Justice 15
   7.1 Simplicity and Class Formation 15
   7.2 Ideal Justice and Other Criteria 16

8 The Implementation Problem 17
   8.1 From Moral Will to Institutions 17
   8.2 The Dark Side of Competition 17
   8.3 Concentrating on “Non-Merit” 18

9 Discussion 18
   9.1 Reconciling Altruism with Reciprocity 18
   9.2 The Problems of Equity Theory 18
   9.3 Monism as Preferable Approach 19
   9.4 The Relation of Distributive and Procedural Justice 19
   9.5 The Social Construction of Merit 19
   9.6 Focusing on the Essentials 19

References 20
1 Introduction

The scientific and philosophical debate about justice has not yet revealed a clear paradigm, with several Theories of Justice competing here (Cohen, 1986; Sandel, 2009; Miller, 2017):

- **Grand Theories of Justice**, usually attributed to some founder (e.g., Sandel, 2009): Virtue – Aristotle, Utilitarianism – Bentham, Autonomy (Deontology) – Kant.

- **Basic principles** that would (alone or in combination) govern the human quest for justice: virtue, happiness, merit, sufficiency, priority, need, equality, liberty, self-ownership etc. While most theories are devoted to a primary principle of justice, some of them make a case for a plurality of principles (Walzer 1983; Audi 2002, see also below on the “tripartite approach”).

- **Different conceptual approaches**, for example naturalism (tracing justice back to natural phenomena – utilitarianism, intuitionism), contractarianism (justice being a kind of social contract), libertarianism (an extreme form of individualism), and several others (cf. Olsaretti, 2018).

We find major conceptual contrasts (tensions, dichotomies). Among these are:

- **Substantive vs. procedural** justice (e.g., Miller, 2017): Can justice strive for certain desired outcomes, or must justice restrict itself to just procedures, accepting any outcome? More generally, social psychologists are concerned with how procedural questions (non-outcome factors) affect perceptions of justice (Lind, 2020).

- **Cognitivist vs. decisionist** approaches (e.g., Quante, 2013, 40ff.): Is justice based on principles to be discovered or (only) subject to contingent agreement? A similar contrast is **rationalist vs. empiricist** (e.g., Binmore, 2005, p. 38): Can we deduce moral principles from reason alone, or do we have to consult data from the real world?

- **The role of intuitions** (Gigerenzer, 2007; Haidt, 2013): Are intuitions (gut feelings) the benchmark for “genuine” justice or merely heuristics for rational reasoning?

This article focuses on distributive justice, i.e., how people allocate benefits and burdens — the most important being income for work. The most general (and commonly accepted) definition is to give each his or her due, a desert-based approach traced back to Aristotle. Distributions based on desert (merit) enjoy practical prominence especially in markets (Miller, 1999; Adriaans et al., 2019; Richters and Siemoneit, 2021). But concerns arise about increasing inequality (Piketty, 2014), and distributive justice remains a major societal and social scientific controversy.

Given this diversity, is a single (let alone simple) Theory of Justice at all possible? Deutsch (1985, p. 35) thought not: justice would be complex and variable. Miller (2017, sec. 7) assumed that “we will need to accept that no comprehensive theory of justice is available to us.” Phenomenologically, Miller (1999) argued that a plurality of three principles would prevail in modern societies: a (moderate) meritocratic principle in combination with a social principle (need) and political equality. This “tripartite approach” emerged in social psychology in opposition to Equity Theory in the 1970s. It quickly and consistently found the support of many authors (among them Deutsch, 1975; Leventhal, 1976a; Lerner, 1977; Mikula, 1980a; Schwindt, 1980; Reis, 1986; Kabanoff, 1991) and today represents one paradigm within the justice discourse of social psychology (Lind, 2020). But the systematic relationship between these principles remains unclear, and with his specific pluralist account Miller would only shift the problem from the principles to the social context (Honneth, 2008).

This article aims to develop a coherent picture of when, why and how people allocate benefits and burdens according to merit, need, and equality, setting it in the evolutionary context of reciprocity. In the terms of Leventhal et al. (1980), it is a theory of allocation preference. It does not discuss non-outcome factors like procedural fairness, for in the end outcomes are decisive. Because there has been a shift from the outcome-focused to the relationship-focused paradigm in social psychology, beginning in the middle of the 1970s (Lind, 2020), many of the works concerning distributive justice cited here date back to the outcome-focused period.

In Section 2 the article depicts the evolutionary roots of reciprocity and how reciprocity translates into the social norm **Meritocratic Principle**. Section 3 explains why one justice dimension is not enough. Moral reasoning is a communicative strategy for resolving conflict, enhancing status, and maintaining cooperation, making justice rather a social bargain and an optimization problem. The picture is therefore extended in Section 4 by social psychological insights regarding intuitions, rules of thumb, and self-bindings. Section 5 and Section 6 show how need and equality can be reconciled with merit. They are auxiliary principles when merit is not effective, not efficient, or not communicable (or simply coincides with equality). An important topic is implementation (Section 8): How is a substantive Theory of Justice “applied”, especially regarding procedural justice? Section 9 discusses the results and formulates a conclusion.
2 Reciprocity as the Foundational Principle of Justice

2.1 The Sociobiology of Cooperation

2.1.1 Selective Forces

Evolutionary selection has shaped a process in which individuals compete for resources to replicate their genes by reproduction, and humans make no exception. Besides competition, fitness can also be enhanced by cooperation, but according to the selection rules every individual must benefit from it in the long run. Today, the existence of genetic altruism can be ruled out — this would contradict any evolutionary functional logic (Voland, 2013, p. 65). Even though human biology does not determine behavior, it limits its flexibility in characteristic ways.

Among animals, mutualism can be observed regularly (behavior that provides direct benefits to every mutualist), often between species (Leigh Jr, 2010). But only higher primates seem to have evolved the cognitive abilities required for reciprocity, i.e., accepting costs for the benefit of others which are rewarded later, either by the beneficiaries themselves (direct reciprocity) or by others (indirect reciprocity) (Trivers, 1971, 2006). Reciprocity is dependent on repeated interactions and it is endangered by cheating, so the cognitive abilities required include individual recognition, temporal discounting, and memory (Stevens et al., 2005). In parallel, mechanisms have evolved for altruistic punishment of cheaters, but also for building up reputation and the assignment of social approval (prestige) (Voland, 2013, 74 ff.).

2.1.2 Forms of Reciprocity

Reciprocity among humans has been widely discussed in anthropology and sociology (see Adloff and Mau, 2005), but only recently in psychology (Kurzban et al., 2015). Note that some authors, especially in economics, restrict the term reciprocity to a personal relation. Reciprocity implies an equivalence in exchange relations: to repay in kind what another has done for us, materially or socially. Voluntary exchange will only occur when the benefits exceed the costs for both partners (costs and benefits in a wide sense, Blau 1968; Becker 1976). Due to different individual levels of marginal utilities and marginal costs both partners can benefit from exchange. This proposition lies at the heart of Social Exchange Theory (Thibaut and Kelley, 1959; Blau, 1964; Homans, 1974), and Neoclassical Economics uses the concept of surplus to characterize such situations.

The condition of equivalence must be met at least in the long run (Kurzban et al., 2015). Sahlins (1965) introduced the concept of generalized reciprocity: transactions that are seemingly “altruistic” but can be expected to be returned, not necessarily here and now, not necessarily by the beneficiaries themselves, undetermined in time, quantity, and value (for a compact overview, see Holcombe, 2020, ch. 5). These transactions might be better described as investments. The modern welfare state, for example, can be regarded as an institutionally mediated, reciprocal arrangement, and several typologies of reciprocal expectations can be construed (Lessenich and Mau, 2005). A decisive point is that free-riding is effectively prevented or punished.

2.1.3 Ideal Justice

From the evolutionary point of view, equivalence is the only stable (and symmetrical) solution to the problem of how two individuals could possibly engage in voluntary social exchange. The idea of equivalence can be readily extended to the negative range: The proportional relation between desert and reward also applies to guilt and atonement (Figure 1). The purpose of punishment is to impose enough costs on the factual or potential defector to offset the temptation to violate the equivalence principle (Boyd and Richerson, 1992; Clutton-Brock and Parker, 1995). Durkheim (1933 [1893], p. 88) emphasized the intended exact balance between the severity of a crime and its punishment (see also Buchanan and Mathieu, 1986, p. 13). People would even “construct” reciprocity if the world is not perceived as just, especially by blaming victims for their “own fault” (“Just World Belief” Lerner, 1980). Austin and Walster (“Equity with the World”, 1975) showed that subjects try to “heal” inequities across several relationships, if necessary.

One of the most striking arguments in favor of reciprocity is the existence of property, money, credit, and prices as such (Holcombe 2020, ch. 5). The idea of property is historically based on the idea of an achievement resulting from one’s own labor (Hume, 1874 [1739–40]; Marx, 1906 [1867]; Locke,
1967 [1698]; Smith, 1981 [1776]), or in a proverb: “as you sow, so shall you reap”. An argument often neglected is the proportionality between the amount of money and quantity of goods. Philosophers of justice regularly argue about the legitimate level of hourly rates (unequal pay, *iustum pretium*, Koch, 1995), but they have a tacit consent that hourly rates as such are legitimate, and so are any other prices. The equation \( \text{amount} = \text{quantity} \times \text{price} \) states that two hours of work will cost two times the amount of one hour (e.g., Tyler et al., 1997, p. 46), reflecting the linear relation of Figure 1.

### 2.1.4 Injustices

The most general definition of a *violation* of reciprocity could be formulated as “advantages at the expense of others”, or “not bearing the costs of one’s own choices”. This holds true for all forms of offense, be it theft, rape, embezzlement, illegal parking, doping or tax evasion. Other terms indicating unbalanced reciprocal expectations are exploitation, recklessness, usury, plagiarism, cartel, extortion, and slavery. In economics, *negative externalities* occur when, due to the activity of another person, one person experiences losses in welfare that are not compensated (Daly and Farley, 2011, p. 184). Under the heading of *rent-seeking*, economists discuss a whole class of activities that attempt to gain larger profits by exploiting or even manipulating economic conditions or public policy, without contributing to genuine wealth creation (e.g., Krueger, 1974). Regulation theory implicitly associates market failures with an imbalance of individual and societal costs and benefits (cf. Shughart, 2008, p. 449). That “free-riding” is a social problem remains unintelligible without assuming a violation of reciprocal expectations (Fehr and Gächter, 1998; Panchanathan and Boyd, 2004).

### 2.1.5 Formulating the Issue

Therefore, the minimum condition of any kind of voluntary social exchange is reciprocity in the sociobiological notion, a ratio of benefits to (opportunity) costs of unity and one that is not too unequal for both, however this is assessed by the individuals – and their collectives (cf. Sandel, 2009, 145f.: “When consent is not enough”). Justice is a social bargain between individuals but witnessed by society.

In *Theories of Justice*, reciprocity is conceptualized as merit, proportionality of contribution, equity, mutuality, or Golden Rule. The theoretical challenge therefore is to integrate the seemingly evident deviations – like need, equality, and “altruistic” acts – into a coherent concept of reciprocity: what could fill the gaps between seemingly higher costs and lower benefits in these cases (cf. Figure 2)?

It is important not to confuse two overlapping, but distinct assessments of social exchange: (1) “Is it just?” This concerns the “public” cost-benefit estimation, i.e., the visible (communicable) aspects of reciprocity that are depicted in Figure 1. (2) “Is it worth the effort?” This concerns the individual, “private” cost-benefit estimation. Beyond the obvious, it includes costs and benefits that are not well communicable or only indirectly related to the transaction and hence unsure: when do benefits cease to be related to a specific social interaction and hence make for “genuine altruism”? The difference becomes salient for example when Nesse (2001b, p. 5) objected to “[t]he tendency to use reciprocity to stand for all cooperative relationships” and listed several “other” reasons why humans would cooperate.

Figure 3 visualizes some of these tensions. While sociobiology concentrates on genetic replication where altruism goes extinct, common sense and parts of the social sciences prefer a semantic dichotomy between “Good and Evil”, restricting reciprocity to a small range of transparent cases (and sometimes to personal relations). The semantics revolve around motives (cf. Homann, 2014, ch. 4): “egoism” is present when one’s own advantages are intended (whether or not at the expense of others), and “altruism” is present when behavior is clearly considerate of others (though personal benefits may arise unintendedly). This article argues that the domain of reciprocity is much broader and that not intentions, but actual costs and benefits should inform the wording. Much semantic egoism is in fact adequate self-regard, and a lot of what is termed altruism should rather be viewed as a “blind investment”. Instead of inventing a new term (cf. the “social semantics controversy”, West et al., 2007, 2008, Wilson, 2008) I will refer to this as “very generalized reciprocity”. True altruism (benefits < costs) as a form of self-exploitation is neither likely nor socially desirable.
2.2 The Meritocratic Principle

2.2.1 History of Discourse

Reciprocity translates into the social norm Meritocratic Principle. The idea of merit is that people deserve certain economic benefits in light of their actions (Feldman and Skow, 2016; Lamont and Favor, 2017). Already Aristotle argued in his Nicomachean Ethics (1131a) that “all men agree that what is just in distribution must be according to merit”.

In Germany, from about 1960 until 1980 three overlapping strands of the debate about “achievement” can be distinguished (citations are by way of example):

- A wage-political struggle between employers and trade unions that continues to this day (performance-related pay) (Pornschlegel et al., 1967; Verband für Arbeitsstudien REFA, 1974; Breisig, 2003),
- a (leftist) critique of a “working society”, “culture of achievement” and alienation of work – and the (conservative) reply (Otte, 1970; Müller, 1974; Jenkis, 1980),
- and how social scientists dealt with the subject (Heckhausen, 1974; Braun, 1977; Hartfiel, 1977; Bolte, 1979).

The debate revived around the year 2000 and culminated in the aftermath of the banking crisis in 2008. While the meritocratic principle was fiercely contested in the 1970s, it was now explicitly demanded (Neckel, 2001; Dröge et al., 2008). Most authors mentioned above agree that merit has many dimensions, and that nearly every aspect of merit (or achievement) can be interpreted positively or negatively. Any objective definition of merit is impossible since it depends on context and circumstances – or the other way round: Merit is uncontroversial only when defined in extremely narrow terms (cf. Deutsch 1985, 133f., Breisig 2003, sec. 3.2). Actually, it is always the beneficiaries of an achievement who will assess its utility. This is a core principle of the division of labor and reflects the general tension between individual and collective perceptions of achievement.

In the UK and the US, the debate had another focus. An important precursor was the so-called Northcote-Trevelyan Report in Victorian England in 1854. Its unanimous recommendation was to organize recruitment and promotion in the British Civil Service exclusively according to merit. This was partly (and correctly) perceived as a critique of aristocratic patronage. In his satirical political book “The Rise of the Meritocracy” Young (1958) coined the term Meritocracy as a dystopia, “portraying a sinister, highly stratified society organised around intelligence testing and educational selection” Dench, 2006, book cover. A collection published after a symposium (Dench, 2006) allows an overview of the debate in the UK and the US.

A proportionality of inputs and outcomes was also the core proposition of Equity Theory which was the leading paradigm of social psychology during the 1960s and 1970s and led to hopes of becoming a general theory of social interaction (Adams, 1963, 1965; Berkowitz and Walster, 1976). Both inputs and outcomes could be positive or negative, and their relation was presented in a complex, somewhat counterintuitive mathematical formula (Walster, Berscheid, et al., 1973). The theory was later criticized for its “unidimensionality” and other conceptual shortcomings and subsequently replaced by the “tripartite approach” with merit, need, and equality (Leventhal, 1980; Schwinger, 1980; Deutsch, 1985; Folger, 1986; Reis, 1986; Lind, 2020).

2.2.2 Definition

The meritocratic principle has many formulations, but often as simple as this: “Those who do more than others shall receive more than others” (Bolte, 1979, 26, translated). Likewise note that also a “reversed equity script” has been reported:
“those who have more must have performed better” (Bierhoff et al., 1986, p. 174). A formal (and lengthy) definition can be found in Neckel et al. (2008, p. 46), but in my view the best definition of merit (especially for modern societies) is the legendary equation of Michael Young in an extended version (Kariya and Dore, 2006, p. 138),

\[
\text{merit} = \text{talent} + \text{effort} + \text{luck}.
\]

More easily than any other definition it can reveal the wrangling over the role of merit. This equation also matches economic theory when interpreting its main terms as “personal production factors”: the natural endowment of individuals with talents (capital) and their efforts (labor), complemented by the contingencies of the economic process and the sometimes hardly predictable assessment by others.

### 2.2.3 Justice Theory as a Social Bargain – Tugging at the Terms of Young’s Equation

Luck is the least contested term in Young’s equation, for it has basically nothing to do with merit. Nevertheless, the term is important for structural and discursive reasons (cf. Dröge et al., 2008, sec. 1.7). First, it is unavoidable, second, people argue about what counts as luck, and third, it is often impossible to clearly separate it from the rest. Luck egalitarianism (e.g., Dworkin, 2000) is effectively a debate about luck vs. choice, i.e., responsibility (cf. White, 2007, ch. 4).

At the least, luck must not be too large. Today, a widespread economic contingency is one of the main acceptancy problems of the meritocratic principle. Neckel (2001, sec. 4) emphasized that merit as a foundation of distribution becomes absurd when jobs themselves have become scarcities (“economic culture of randomness”). On the other hand, Saunders (2006, p. 192) argued that luck cannot be systematically exploited, and everyone can have it, so it would not question the legitimacy of meritocracy.

Talent (or ability) is ambiguous. Exemplary for many, Rawls (1999, p. 86) emphasized that “inequalities of birth and natural endowment are undeserved”, results of a “natural lottery” (p. 64). At the most, achievements based on talent could entitle legitimate expectations from society, given that appropriate rules had been established in advance: “But this sense of desert is that of entitlement” (p. 88f.). Rather like in a tug-of-war, Rawls tried to “tug” talent toward the term luck and to turn a moral desert into a question of contingent entitlement. Libertarians argued the other way round (cf. Sandel, 2009, p. 69): Talent requires a lot of effort to develop (“knowledge”), and who else but the talented should be allowed to benefit from his or her talent? Libertarians tried to tug talent toward the term effort and turn a moral desert into a question of property rights. The resemblance to the nature-nurture debate is obvious.

The term effort is indeed for many the allegory of a deserved income. Everywhere in the world hard working people are acknowledged as creators of economic value – at least if this effort is targeted at a socially desirable outcome and achieves a result. Throughout the debate the tension between effort and achievement is a recurring topic: effort without achievement is a pity, achievement without effort has no merit (Heckhausen, 1974; Braun, 1977; Bolte, 1979; Deutsch, 1985; Neckel, 2001; Dröge et al., 2008). Neither entitles reward, and it could even be argued that the equation should rather read merit = talent × effort + luck, so that if either talent or effort are zero, their product would be zero as well, the remainder being luck. Yet the basic problem of “performance-related pay” is that while its normative basis can only be merit, its transparent (communicable) basis can only be achievements (Breisig, 2003, sec. 2.2).

Rawls (1999, sec. 12) went even further and declared that the ability to make an effort was dependent upon a happy childhood and social circumstances (cf. also Sandel, 2009, p. 158). In fact, Rawls tugged any term of Young’s equation toward luck, apparently to be free for his contractarian approach to justice, i.e., contingent agreement (cf. Miller, 1999, pp. 53-59).

Karl Marx concentrated on effort and called this, following his classical antecessors, a labor theory of value. In his writings Marx spoke of the equality of human work, of the time elapsed as a measure of effort (Heinrich, 2005, 71ff.), but also that a higher qualification would cause higher costs, and hence justify a higher value of the working time (hourly rate, p. 91). He largely ignored talent and fully disregarded the contribution of a whole class, namely capital as the merit of the entrepreneurs, which he assumed to be mere organization plus supervision of exploitation and therefore zero (p. 157).

Talent and effort have yet another dimension: Societal functionality. Early accounts of the social benefits resulting from functional stratification are Durkheim (1933 [1893]) and Davis and Moore (1945). The importance of not discouraging the talented is another recurring topic of the debate about merit, tightly connected to secure property rights (cf. Acemoglu and Robinson, 2012).

### 2.2.4 Conclusion

Contrary to Young’s apprehension and other objections, the meritocratic principle is a fundamental and widely accepted social norm: Meritocracy “resonates powerfully with deeply held ethical values about fairness, and these are broadly shared throughout the population” (Saunders, 2006, 193, original emphasis). It “corresponds to the widespread belief that people deserve to enjoy unequal incomes depending on their abilities and how hard they work” (Miller, 1999, p. 178). The meritocratic principle establishes a relationship between personal
market value and contribution to productivity (Marris, 2006, p. 159).

Unlike many other theorists, Miller (1999) considered popular conceptions of justice (“what the people think”) and empirical data. He stated that desert and need criteria feature prominently here (p. 90), and criticized political philosophers, Rawls among them, for neglecting such empirical evidence, especially when it is in favor of desert (ch. 7). Further empirical support for merit as a base for distributive justice can be found in Saunders (2006) and Neckel et al. (2008). Recently, Adriaans et al. (2019) showed that there is widespread approval in Germany and the other European countries for income distribution according to equity and need.

The results of the scientific and philosophical debates about the meritocratic principle can be summarized as follows:

- It is defendable as a theoretical base of justice.
- It is functional in increasing societal wealth.
- It is not possible to define merit objectively.
- In practice, the principle is often not met (effortless incomes, underpaid work, the competing principles need and equality).

Most objections against the meritocratic principle can be rebutted by concentrating on “non-merit” (Section 8).

3 Justice as a Social Bargain and Optimization Problem

Figure 1 depicts the ideal of justice, demanding to appreciate the circumstances (costs and benefits) of every single case. But neither our physical nor our social world are ideal. Since justice is a discursive, nonviolent concept (like arguments in general, Kopperschmidt, 2000), we can at least set aside questions of power and force for a Theory of Justice. Then still among the real-world restrictions are (1) costs, time constraints and scope for information retrieval and communication, (2) general uncertainty about future events, (3) strategic behavior of our fellow humans.

Moral reasoning is a communicative strategy for resolving conflict, enhancing status, and maintaining cooperation (cf. Deutsch 1985, p. 26, Mikula and Wenzel 2000, Haidt 2013), making justice rather a social bargain and an optimization problem. It includes assessments, compromises, investments, renunciation – and above all a pool of shared “good reasons” (plausibility resources) as an argumentative basis (Kopperschmidt, 2000).

The ideal of justice is just one of many dimensions (cf. Reis 1984, Deutsch 1985, p. 99, Young 1994), and not always the strongest force. In practice, justice must in my view fulfill at least three additional requirements: effectiveness, efficiency, and communicability. For each the Aristotelian “mean between extremes” is optimal, but since trade-offs exist between them, we actually have an optimization problem.

Effectiveness: functionalist ~ expedient ~ sticking to principles

Justice must be useful, that is the quintessence of utilitarianism. If justice degenerates into “pure principle” (dogmatism), creating problems rather than solving them, it misses the target (e.g., Kant’s murderer dilemma, Sandel 2009, p. 132). Mikula (1980a) discusses extensively the role of “allocation goals”, and Greenberg (1982) and Leventhal (1976b) are rich sources of pragmatic considerations when choosing an allocation rule. Effectiveness has priority when it comes to rationing (Elster, 1992; Young, 1994). On the other hand, justice must not be functionalist, that is the quintessence of deontology. If justice uncritically maximizes “outcome” (especially in the short run) it becomes arbitrary, and many individual and societal self-bindings remain unintelligible (Sandel, 2009, 32f.).

Efficiency: sweeping ~ practicable ~ sophisticated

Since humans have evolved to maximize efficiency (Sanderson, 2001, p. 148), justice should be rather straightforward and “easy”, at least practicable. It must be applicable to everyday situations, not only to philosophical thought experiments. Long theories, complex reasoning and the elaborate dissection of moral subtleties are as “unjust” as endless meters of tax legislation and its commentaries. But justice cannot be allowed to simply flatten all relevant differences between the cases in question in order to push through one single principle. We cannot deviate too far from our ideal of justice to appreciate the individual circumstances of every single case.

Communicability: populist ~ thoughtful ~ unworldly

Justice must be “reasonable”. A Theory of Justice must be consistent and withstand rational reflection. Any distribution rule must fulfill the rule of the good reason (Section 4.3) and “treat like cases alike”. Armchair theories that contradict basic intuitions of justice and ignore empirical evidence are to be refuted (Miller, 1999). On the other hand, justice cannot be merely a question of majority (vox populi) or strong emotions. A precondition of a shared sense of justice is comprehensibility and transparency of the criteria. Justice must be socially monitorable.

To derive the proper consequences for a Theory of Justice we must extend the picture, because we need to know more about the social psychology of humans regarding (moral) intuitions, the role of rules of thumb, good reasons, and simplicity in efficient decision making, and – especially important – self-bindings as a precautionary device and “blind investment”.

---

Siemoneit: Justice as a Social Bargain and Optimization Problem
4 Extending the Picture

4.1 Self-Binding

4.1.1 Theorists of Self-Binding

Self-binding (also called commitment or precommitment) was made a topic in economics by Thomas C. Schelling (1960). He regarded conflict situations between several parties rather as bargaining situations, and believed that the public and credible renunciation of realistic options could improve one’s bargaining position. This was in stark contrast to the neoclassical assumption that more options are always better. “Burning one’s bridges behind oneself” to unmistakably make clear that retreat is no option, or “calculated madness” (brinkmanship) can force the other parties to make a complete reassessment of the whole situation. Schelling later extended the idea of conflict to intra-personal “two-mindedness” (Schelling, 1978, 1984, 1985).

The idea of commitment can be readily extended from conflict situations to social behavior in general, a topic of Jon Elster’s (e.g., 1979). A self-binding is to act now to ensure a future act that could but would not have been performed without that prior act (Elster, 2003, 1754, discussing constitutions as collective self-bindings). Since by definition trade-offs exist, any self-binding is endangered by a change of mind, therefore in Elster’s opinion a self-binding must be public, and its revocation must be costly (or impossible). Elster is very strict about two assumptions: (1) The outcome must be intended. Any “unconscious” self-binding would devalue it. (2) Only individuals can bind themselves. Any notion of collective self-binding (institutions) is to be explained in terms of individual (and intentional) self-binding. Elster makes further comparisons between individual commitment and constitutions: Despite structural similarities, an individual would lack the possibility of a separation of powers, and attaching costs to some options would be no feature of constitutions (Elster, 2003, 1754f.).

For Karl Homann (2003, part B), self-binding is a means for more freedom. Self-bindings make individuals reliable by removing unfair options from their repertoire of actions, and this reliability of mutual expectations enables new and beneficial cooperation between self-interested individuals (p. 73 and ch. 6). Homann is a critic of “too much” individual self-binding (decency, appeals to values, Christian ethics etc.), and he accuses philosophical ethics of being heavily biased toward the idea of a “genuinely moral motive” (cf. Section 8). Most modern problems would instead be the non-intended side-effects of “normal” behavior, so any higher morality would be no remedy, especially because of the exploitability by others. Instead, he emphasizes the significance of institutions as collective self-bindings.

4.1.2 Properties and Examples of Self-Bindings

Self-bindings are part of the “little tricks we play on ourselves” (Schelling, 1978, p. 290) to limit our potential to dupe ourselves. This is a classic case of utility maximization in the long run (the time scale considered is crucial in this respect). For the topics discussed here I suggest the following definition:

A self-binding is the strict adherence to a contingent rule which aims at some “higher goal” that is difficult to achieve directly.

I prefer the term self-binding over commitment because the latter is too near to a (personal) promise and lacks the contingency aspect. A self-binding lies somewhere between “pure investment” where costs now will surely reap benefits later, and “pure superstition” where costs now definitely have no connection to assumed benefits later. “Higher goal” means that the connection between the rule (costs) and the desired outcome (benefits) is not transparent. Their probabilistic relation may be plausible in some cases but controversial in others (see examples below). Therefore, the rule is contingent. But the rule must be strictly adhered to for it to be communicable, fulfillable (clear criteria), and socially monitorable. Hence, a self-binding could also be viewed as a “blind investment”. In my view it is speculations about the long-term significance (“sense”) of certain self-bindings (“values”) and the individual (non)willingness to make such blind investments that make societal debates so wide-ranging (and indeed heated).

I would rather give up several of Elster’s restrictions:

- There is no need to restrict self-bindings to be intentional (conscious). Many interesting self-bindings would not pass this test, and even Elster struggles with that somewhat contingent distinction (e.g., Elster, 2003, p. 1761). The unconscious is our unconscious that pursues our goals and can make judgements on its own (Sumser, 2016, 275, original emphasis).

- A self-binding does not need to be public. Any intent (e.g., not becoming addicted to alcohol) that is safeguarded by a contingent rule (“no alcohol before 8 p.m.”) exhibits the characteristics of a self-binding, a view supported by Schelling throughout his contributions.

- Contrary to Elster, personal self-bindings could be effective because individuals have a separation of powers, identifying the roles of individual action, moral standards, and conscience (guilt, shame) with the executive, legislative and judiciary, as elements of intrapersonal communication (cf. the theory of self-command, Schelling, 1978).

Self-bindings can be applied on different levels, from the individual up to the global community, and their characteristics
depend on goal, criticality (what is at stake) and transparency (communicability). A structural characteristic is that a self-binding on a higher level (collective) is a coercion on the lower level (the collective’s individuals) when a self-binding on the individual level seems rather unpromising.

The following list of examples is neither exhaustive nor strict in its categories (see also Nesse, 2001a, for an overview):

- **On the individual level**, we find concepts like civility and good conduct, moral principles, loyalty, veganism or a “pro-life” stance, awareness-building, religious dietary laws, superstition, and political correctness. People believe rather than know that certain behavior would be favorable or adverse for themselves or for running a society, and they discuss passionately about it (Haidt, 2013). Self-bindings can be part of a (political) identity and contribute to authenticity (e.g., Greenebaum, 2012). The higher goal of credibility may sometimes require even “silly” investments keeping a ruinous promise, carrying out a senseless threat, Nesse, 2001b, p. 12.

- **On the level of societal subsystems**, we have most prominently science and the judiciary. For the higher goal of finding “objective truth” in a society of strategic self-interested individuals, both have given their systems strict, transparent rules so as to avoid or correct any biases of their members and maintain a high standard of impartiality. These rules are continuously reflected upon.

- **On the level of societies** (nation states), we observe restrictions regarding the ownership of weapons or consumption of drugs. During the Corona pandemic many legislations have created a host of contingent, quickly changing and hence contested rules regarding personal contacts. Constitutions are among the highest-level national self-bindings we know, “intricate pieces of machinery” (Elster, 2003, p. 1779). Between nations we find concepts like nuclear deterrence or Swiss and Swedish neutrality.

- **On the level of God**, the existence or lack of self-binding has been discussed since time immemorial (“Theodicy”).

Self-bindings can be nested (several levels of judicial appeal), staggered (adherence to a self-binding is safeguarded by another self-binding), and distributed (network of mutual control, like the “scientific community” or separation of powers).

The tension between rigidity and flexibility of self-bindings is a recurrent topic. The “dogmatic” adherence to a self-binding is in fact its caricature (because the rule itself is contingent), but a “flexible” self-binding is neither reliable nor communicable or monitorable, so at least public collective self-bindings can be nothing but “rigid”. Handling self-bindings inconsistently foments a suspicion of bias. Practically, when a self-binding becomes absurd (“Fiat iustitia et pereat mundus”) or when all people concerned share a strong interest in not following the rule, they usually search for justifications to discard it or to go to the limits of latitude.

### 4.2 Moral Intuitions

Gut feelings (intuitions, hunches) refer to a judgement (1) that appears quickly in our consciousness, (2) whose underlying reasons we are not fully aware of, and (3) that is strong enough to act upon (Gigerenzer, 2007, 16ff.). Intuitions work by using simple rules of thumb (heuristics), which take advantage of evolved capacities of the brain. Although intuitions often use very little information, they can be amazingly accurate and better (or at least not worse) than much more complex strategies (ch. 8). The “intelligence of the unconscious” lies in choosing a rule of thumb that is appropriate for the situation and relying only on the most relevant information (“the essentials”) while neglecting the rest.

Humans are social, and humankind has evolved much of their time in groups of manageable size. The complementarity of individual development and group stability requires a delicate balance (or trade-off) between striving for autonomy and accepting heteronomy and hierarchy. Moral intuitions assist in finding that individual balance (cf. Sumser, 2016, sec. 4.2). Although moral rules seem to hold universally, the subject experiences the source of its moral judgement in itself, connected with sentiments of obligation and commitment. Regarding morals, everyone is an expert.

In the last two decades, two models developed for understanding moral judgement have garnered prominence. Haidt (2001) presented the “Social Intuitionist Model” (SIM) which initiated a turn from rationalism to intuitionism in moral psychology (cf. also Haidt, 2013). According to the SIM, rational reflection is neither the cause nor even the starting point of a moral judgement. Instead, “intuitions come first, strategic reasoning second” (Haidt, 2013, p. xiv). Moral intuitions have a primacy, they allow an immediate and effortless assessment of moral situations. The post hoc reasoning has above all a communicative function of (social) justification. Reasons do not have to be true but rather plausible and acceptable (“press secretary, not independent judge”). Intuitions are not immune to the reflection of “thoughtful and well-educated” people (Ross, 1930, p. 40), but to change one’s mind they usually need to activate another intuition. For justice, especially significant is the tension between a strongly felt moral conviction and its bad communicability (“moral dumbfounding”, Haidt, 2013, ch. 2).

The SIM was challenged by “Triune Ethics Theory” (TET, Narvaez 2008, 2010). According to TET, reflection plays a much larger role than suggested by the SIM. Reflection can be
made possible by conscious deliberation, moral development, and developing expertise. These abilities can be somatized again and then “look like intuition” (Narvaez, 2010, p. 167).

Both models can be viewed as quite complementary rather than contrary, the SIM describing more everyday situations, the TET more “professional” moral judgements. This way or the other, moral considerations begin with an intuition, may be consciously reflected if necessary, and become a judgement only when it feels emotionally safe (Sumser, 2016, sec. 8.1). Social framing can strongly influence moral judgements, as can priming emotional flashes like disgust, Haidt, 2013.

4.3 Efficient Decision-Making

4.3.1 (No) Good Reason

According to Gigerenzer (2007, ch. 8 and references therein), our intuitive judgements are often based on only one good reason. This is contrary to the tenets of rational decision theory which holds that we must consider and weigh all relevant information. A reason is a cue or signal that is significant for a relevant fact, and humans use reasons to make decisions intuitively and efficiently. The Take-the-Best heuristic is an application of that principle: reasons are ordered by relevance and then checked successively (lexicographically) as to whether they enable a decision. Instead of weighing all options “until the end”, the most relevant reasons are checked first.

4.3.2 Rules of Thumb

A heuristic is a problem-solving strategy for quickly finding a satisfactory solution at the expense of accuracy, especially when the problem is complex or when information is scarce. A rule of thumb can be defined as a default heuristic for a specific problem that is regularly encountered. Rules of thumb are applied in bodily movements (e.g., catching a flying ball) as well as in mental processes. They play an important role in moral behavior, the paradigmatic example being the Golden Rule: “Treat others as you would like others to treat you”. Another prominent example for a rule of thumb is “tit for tat”, a simple recipe for successful cooperation without being exploited (Axelrod and Hamilton, 1981): Start with cooperation and then replicate every opponent’s move (cooperation or defection). Both rules are strictly reciprocal.

Gigerenzer (2007, ch. 10) states two examples where rules of thumb can be viewed as social instincts: (1) Strong group cohesion supports conformity to group norms and hierarchy, and the rule “don’t break ranks” can explain cases of participation even in barbaric acts against one’s own convictions, as for example in war. (2) Many jurisdictions require potential donors of organs to consent in advance, but the degree of consent depends heavily on the design of the questionnaire (nudging): Most people choose the “default option”, i.e., the option that for some reason is more straightforward than the others or even pre-selected (“focal point”, Schelling, 1960). The rule of thumb reads: “If there is a default, do nothing about it”. Another example given for social instincts is imitation (Gigerenzer, 2007, ch. 11): “Do what the majority of your peers do” is a good recipe for avoiding failure, but it can be overruled by “Do what the successful do”.

Rules of thumb are applied quickly, effortlessly, and usually unconsciously, but can be – as suggested by SIM and TET – modified or overruled by reflection, though this is rare. Another rule of thumb especially important for a Theory of Justice – equality as default – is discussed in Section 6.2.

5 Need as Auxiliary Principle

5.1 The Domain of Need

Of all principles, need seems to be most apt to explain “altruism”: people not able to maintain themselves receive aid from others who cannot expect any reciprocal return (“giving something for nothing”, Gouldner, 1973). But Kurzban et al. (2015, p. 582) pointed to the fact that the costs for any aid for non-kin must be recovered at least on average, otherwise this behavior could not have been selected for in the evolutionary process – hence, it is not altruistic.

This section discusses two topics: (1) Exchange in the family is governed by reciprocity, not need. (2) Institutionalized aid to non-kin (welfare) is dependent on many preconditions, the most important being “deservingsness”, what can be best interpreted in terms of reciprocal expectations.

5.2 Need in the Family?

Frequently, families are stated as a counter example to reciprocity, since children would not be able to reciprocate (Gouldner 1973, Tyler et al. 1997, p. 51, Heidenreich 2011, p. 130), or families would obviously distribute mainly according to equality and/or need (Lerner et al. 1976, p. 153, Hochschild 1981, p. 107, Deutsch 1985, pp. 29f., 42f., Miller 1999, p. 26). Viewed sociobiologically, it is evident that kin selection rules family life, but a reciprocal vocabulary can also be used.

The basic “utility” of children lies in replicating their parents’ genes, and evolution could hardly select for anything else. Walster, Walster, et al. (1978, ch. 6) discuss in depth the costs and rewards of the parent-child relationship in terms of Equity Theory, and they cite evidence that a kind of “immortality” is a strong source of parental satisfaction. It is part of the parental dilemma that children accomplish this “achievement” simply by birth, and from that moment on their genetic interests differ from that of their parents (Kurzban et al., 2015). These genetic
parent-child conflicts become evident for example in weaning conflicts and cry-babies (Voland, 2013, sec. 4.6). If children do not “cooperate”, parents cannot withdraw their support without risking their parental investment, and parental love has evolved to mitigate this dilemma (Trivers, 1972; Daly and Wilson, 1999).

This “economic” or “biologicstic” wording may sound repulsive (cf. Graeber, 2011, 91ff., wrestling with this question), but sociobiology provides overwhelming evidence that it is the biological descendancy that is decisive for parental investment, not the “family” social unit (for the following cf. Voland 2013, ch. 3 & 4, with even more examples, or Daly and Wilson 1999):

- Frustrated reproductive expectations are probably a main reason for divorces, and in many cultures any “constraints” on reproductive performance are accepted as reasons for divorce. Stepchildren, unsure biological fatherhood, and sick or disabled children are destabilizing factors for marriage.

- Stepchildren or children with unsure biological fatherhood suffer massively higher risks of accident and death, have higher stress levels and are often materially disadvantaged, compared to biological children.

- Parents try to massively intervene in decisions of their children, ignoring their “needs”, e.g., with mating prohibitions (“Romeo and Juliet”) or appeals to a “family solidarity”.

Although “Life Course Reciprocity” plays an important role for both directions of intergenerational exchange (Silverstein et al., 2002), parents do more for their children than vice versa during their lifespan, even when not considering childhood and youth phase (Hollstein, 2005, 196ff.). All this makes clear that in the family “need” is rather a small wave on a deep lake of (hardly quantifiable) reciprocity, and many (grand) parental responses to the needs of their (grand) children can be rightfully regarded as investments.

5.3 The Role of Deservingness for Welfare

Modern societies have usually institutionalized aid to non-kin as “welfare”, with full social inclusion as one of the central aims of social policy (see Oorschot and Roosma, 2017, for an overview). This is not self-evident since there is “a social norm against living off other people and a corresponding normative pressure to earn one’s income from work” (Elster, 1989, p. 101).

During the last decades there has been a growing body of literature on the role of “desert” in social welfare, because “the deservingness opinions of various social actors play a pivotal role in the social legitimacy of welfare schemes” (Oorschot and Roosma, 2017, p. 4). Especially important is the repeated finding of a “universal dimension of support” for certain welfare schemes: “[T]he rank order of the average deservingness of the groups of ‘the elderly’, ‘the sick and disabled’, ‘the unemployed’ and ‘immigrants’ tends to be the same” in all European countries (cf. Oorschot and Roosma, 2017, 14, 20ff., and references therein). The basic conditions for the legitimacy of welfare are that (1) aid as such is restricted to those who “deserve” it, and (2) the extent of aid is restricted to “need”, i.e., objective and legitimate necessities.

Oorschot (2000) has developed a framework of five criteria (the “CARIN” scheme) that can largely explain differences of the perceived legitimacy of targeting aid to target groups. Even though reciprocity is among the criteria, this is “obvious” reciprocity. But all criteria can serve as proxies for justifying a “very generalized reciprocity” on the part of the donors, and they match with the ideal of justice. They exhibit the properties of good reasons and are well communicable.

- Control: Fate or fault? If the recipients are personally responsible for their predicament, they are less deserving (if at all). Regarding unemployment, an important parameter is the availability of jobs: A high unemployment rate reduces the personal responsibility for being jobless (Fridberg and Ploug, 2000; Jeene et al., 2014). My interpretation: Violating the norm of personal responsibility reveals a lack of “reciprocal precaution” – or in other words: guilt, and the ideal of justice requires some atonement to prevent thoughtlessness from being an attractive option.

- Attitude: Eager or sloppy? Signs of compliance to welfare conditions and to expectations of the donors in general (e.g., gratitude) increase deservingness. Kootstra (2016) found that receivers of welfare who have a long work history and invest great efforts into finding a new job are considered more deserving. My interpretation: These signs are proxies for social reliability, a will to engage in one’s own contributions and not exploit the situation.

- Reciprocity: Gift or compensation? During the need situation, reciprocity cannot be expected, but already acquired merits (work history) as well as merits likely in the future increase deservingness, for obvious reasons.

- Identity: Us or them? The “closer” the recipients are to the donors, the higher the perceived deservingness is. My interpretation: The paradigmatic example of closeness regarding welfare is “nation”, and Miller (1995) provided an account of how a “well understood” sense of nationality can contribute to adhering to a nation as an ethic
community. A functioning nation is a good precondition for a high level of generalized reciprocity, as opposed to mere fairness, since generalized reciprocity affords a limited group (p. 70ff.).

- Need: Basic or superfluous? Greater need means more deservingness. My interpretation: It is always the donors who define what is considered as need, and they restrict aid to what is necessary, i.e., “true” or “objective” needs. Any kind of institutionalized welfare aid has relatively low upper limits (“socio-cultural breadline”), while the avoidance of misery specifies the lower limit of aid. Extensive checks of available means of the needy are common. This can be understood as limiting the deviations from reciprocity to an acceptable extent.

5.4 Summary
The notion of desert governs the domain of need, and desert is used to justify reciprocal expectations and deviations from it. Need is no desert, but it can (and should) be a case for entitlement. Need has already been interpreted as a metaphor for a Mutual Insurance Society (Lucas, 1972), the main fear being free-riding, so several precautions and limitations are necessary. Acts of welfare can be interpreted as a “very generalized reciprocity”, an investment of people into the cohesion of society – their society.

6 Equality as Auxiliary Principle

6.1 The Domain of Equality
As a principle of distributive justice, equality is much more demanding than need since need is quite well-defined and has a clear scope. Equality is less specific, and especially for distributing income it is of little empirical relevance (see below).

This section discusses four topics: (1) Equality can be efficient as a default (or fallback) when lacking good reasons to choose otherwise. (2) Equality of material rewards for individual achievements (income) is restricted to groups of manageable size and can be reconciled with reciprocity. (3) The access to a decision about the distribution of (dis)advantages is equal for all (equality of access). But the decision itself is basically about merit or guilt, and equality of access can be best interpreted as a self-binding. (4) Political equality is primarily the lack of good reasons for inequality (and a self-binding).

6.2 Equality as Default
Equality can be a fallback when the relevance of other criteria is not given or not known precisely enough. “With a good reason, accept inequality – with no good reason, strive for equality” is a widely cited rule of thumb (Wollheim and Berlin 1956, p. 305, Frankena 1962, p. 5, Benn and Peters 1965, p. 128; indirectly Schelling 1960, p. 65, Mikula 1980b, p. 140, Buchanan and Mathieu 1986, p. 15, Elster 1992, p. 70, Young 1994, p. 8, Miller 1999, p. 233, White 2007, p. 11, Heidenreich 2011, p. 177; experimentally Deutsch 1985, ch. 11; dissenting: Kolm 1996, p. 37; for even more references see Gosepath 2011, sec. 2.4). Generally, justice theory can be stated as justifying deviations from equality (Elster, 1992, 200, original emphasis).

When lacking a good reason, equality is always communicable and difficult to challenge, it “fills the vacuum of indeterminacy” (Schelling, 1960, p. 73). An equal distribution has the additional advantage of not requiring any effort to assess individual contributions and to defend the final distribution (Mikula, 1980b, p. 131) – equality is easy, and Walster, Walster, et al. (1978, p. 213) provide some efficiency considerations regarding “When Equality? When Proportionality?”: time constraints, communication costs, the value in dispute, or the significance (precedence) for future decisions. But regarding communicability, even the proportionality rule is “equality”, albeit of units of claim (Young, 1994, p. 80).

6.3 Equality in Groups
6.3.1 Overview
Since the 1970s, and in contrast to Equity Theory, several social psychologists have developed multi-dimensional approaches to justice. They connected distributive principles with types of social relationships in groups and their specific goals, characterized by a tension between equality and equity (cf. Druckman and Wagner, 2016, 389f.). For example, Deutsch (1985) contrasted an egalitarian, solidarity-oriented group (focus on social relations) with a meritocratic, economic-oriented system (focus on individual outcomes). Kabanoff (1991) observed equality combined with an emphasis on solidarity, while equity (proportionality) was combined with an emphasis on productivity.

Solidarity should not be confused with the principle of need. Solidarity is a reciprocal relation with a norm of mutual obligation to contribute one’s share to a common goal (Bayertz, 1999). Group dynamics show the interdependence of group cohesiveness, performance, and task commitment (Mullen and Copper, 1994; Porter and Lilly, 1996).

6.3.2 Confluence of Equity and Equality
Distributions based on merit can coincide with those based on equality when everyone’s contribution (input) to a common output is of equal value (either actually or supposedly), which Leventhal (1976b, p. 115) called the “confluence of equity
and equality norms”. But then the principle of distribution is still merit, not equality. Investigating “backward reasoning” from visible rewards to unknown performance, Bierhoff et al. (1986, p. 181) reported that unequal rewards would activate the equity script, whereas equal rewards would activate both the equity script and the equity script, not the equality script alone.

There is a good reason to choose an equal distribution – equality is not a fallback here.

6.3.3 The Communicability of the Chosen Distribution

But it would be a mistake, however, to contrast these two allocation norms too sharply. Leventhal (1976b) discussed in detail the role of an allocator in groups or organizations. Allocators would usually choose a compromise satisfying both norms partially (p. 114). They must balance several dimensions, among them the productivity of the group, avoidance of conflict, and their own approval and acceptance by the group members (authority).

The chosen allocation rule is strongly contingent on communicability considerations: when allocators can conceal their allocation decision to the group members (secrecy), they are inclined to follow the equity norm, for reasons of justice, productivity, and reduced group conflict due to a lack of social comparison (Leventhal, 1976b, p. 111). When group members are strongly interdependent, when they perceive themselves as “similar”, when their success can only be credited to the whole group, and when group productivity can be fostered by keeping up a “solidary spirit” within the group, several authors reported a preference for an equal distribution (Leventhal 1976b, 108ff., Greenberg 1982, 418ff., Deutsch 1985, p. 147).

It seems that an individual’s contribution to a genuine group achievement must stand out conspicuously to justify and enforce an unequal distribution of rewards.

6.3.4 The Communicability of Costs and Benefits

Social exchange includes tangible goods or services, but also less “concrete” resources like status or information (Resource Theory, Foa and Foa, 2012). Less concrete does not mean less valuable or not targeted at material gain (which from the sociobiological point of view is the decisive outcome). Frank (2000, ch. 9) vividly described that status gain is not an objective in itself, but functional: A means to achieve higher income, better health, or more promising social relations (esp. in mating).

Long-term cooperation is contingent on many circumstances, so it can be advantageous to forego immediate benefits. Mikula (1980a) and Schwinger (1980) reported that in groups a “politeness ritual” would prevail when the group members themselves were to propose a distribution rule: High performers tended toward an equality rule and low performers toward a proportionality rule, both seemingly at odds with their material interests. Yet Bierhoff et al. (1986) cited several studies where high performers tended toward more equal distributions if future interactions could be expected (but to equity otherwise). This can be regarded as a self-binding (investment) since reciprocity by definition depends on repeated interaction. Social harmony, less conflict, and a high degree of group cohesion are less visible benefits, justifying very visible costs.

6.3.5 The Limits of Performance-Related Pay

Deutsch (1985) markedly criticized a lack of empirical evidence that, within a group, individual productivity would be increased by equitable rewards, making a case for equality as a default reward system in groups. Indeed, ‘Performance Related Pay’ as the periodic adjustment of wages and salaries to individual performance is ever since plagued with countless difficulties, especially when job responsibilities are ambiguous (e.g., Isaac, 2001, Perry et al., 2009, Breisig, 2003). As discussed in Section 2.2, merit is uncontroversial only when it can be related to objective criteria.

But Deutsch underestimated the significance of performance regarding routinely paid wages and salaries. His experiments lack an ‘employer’ who has a personal interest in high performance at reasonable costs and who competes for the best and brightest with other employers. Unequal pay according to qualification and performance is general practice, and Greenberg (1982), Leventhal (1976b) as well as Walster, Walster, et al. (1978) provide many of the issues employers would consider when choosing levels of pay. Performance related pay in this sense is used to find (and keep!) the right people for the right job. The interesting feature of a market price is that it is a collective, rather neutral assessment of performance and achievement, a condition usually not met in small groups.

6.3.6 The Case of Israeli Kibbutzim

Kibbutzim are voluntary rural communities in Israel that share an egalitarian ideal. They are characterized by communal living with equal sharing of outcomes, no private property, and no cash exchange inside the kibbutz. The main rationale for the existence of kibbutzim is economic risk-sharing, average group size is about 440 (for this and the following cf. Abramitzky, 2011).

Kibbutzim largely face three threats, as predicted by economic theory and empirically verified: the exit of more productive members (brain-drain), the entry of less productive members (adverse selection), and shirking in the communal work (free-riding). In effect, kibbutzim must set straight the reciprocal calculation for each member, be it a high or low
performer. This includes incentives and sanctions, some of them nonmonetary, feasible only in small groups: tight personal control, social prestige etc. Kibbutz members must be constantly reminded, by education and ideology, that they are investing into a common project. While kibbutz members see themselves as counter examples to economic man, Abramitzky showed that economic principles do not end at the kibbutz gate.

6.4 Equality of Access

Application processes are generally characterized by an equality of access. It would contradict the idea of an “application” when applicants with a prima facie claim on the good being allocated are treated differently before the decision process (cf. Miller, 1999, ch. 5). But the decision is basically about merit or guilt.

Equality before the law primarily affects the ability to be indicted, the possibility to “apply” for a trial, no matter what status or “deserts” one has. During the trial, procedural justice requires all defendants to be treated impartially, but substantial justice requires the opposite. The exact purpose of a trial is to determine objectively the extent of guilt and to find out what the just deserts are.

Likewise with political offices. The goal of every election of representatives is to bring only the best and brightest in offices, not anyone. But we admit everyone to be elected, again for reasons of self-binding, and merit is assessed afterwards according to results. If in doubt the person will be unelected on the next occasion. But if there is a candidate of whom we know that she is the best, she will certainly win the election, yet not according to criteria determined in advance, but based on a collective assessment.

The same holds true for all other applications, be it a new job or a place at university – equality is equality of opportunity, the rest is about merit, but this will be assessed afterwards. It is a common misunderstanding made in philosophical debates on justice when asking “Who deserves a job?” (cf. Miller, 1999, ch. 8: “Deserving Jobs”): already the hiring is regarded as a reward. But the rewards only come later, as monthly payment for achievements or as high grades for good exams, i.e., according to merit. The successful application as a chance of a true test can be viewed as a reward in a factual sense, but not in the systematic sense discussed here.

6.5 Political Equality

Suffrage or membership in political bodies has become ever more inclusive because ever more reasons for inequality revealed as dysfunctional in an ever more liberal society had to be dropped. In fact, no reason has remained except a (contested) minimum age, and the self-binding of equal dignity and respect for all humans is now widely accepted as a minimal standard (cf. Gosepath, 2011, sec. 2.3). Most societies would profit if only “thoughtful and well-educated” people could vote (noocracy), yet try to operationalize that. Already the idea seems absurd, and all respective historical attempts have failed spectacularly. Especially when placing people in positions where they are to expend other people’s money (e.g., taxes for public expenditure) or decide over access to advantages we must be very cautious with conflicting interests. It could put the fox in charge of the henhouse, though, when giving the wealthy a higher share of votes, and the same holds true for all other reasons once deemed good. “One person, one vote” therefore is not rooted in the idea of equality as such, but in an effective self-binding on a highly sensitive field.

6.6 Summary

Equality is not a principle of distributive justice but a fallback, chosen for a lack of good reasons to choose otherwise. Like need, equality can be a case for entitlement. Of course, it is contested what good reasons are (or are not), but for many important cases they are uncontroversial. Some reasons once good have lost their plausibility and have been removed from the “societal pool” of arguments. Other cases often stated as examples of “genuine” equality have revealed their reciprocal nature and their compatibility with utility maximization, though sometimes only as a cost avoidance strategy.

Inequality is often identified with unfairness, but the opposite seems to be general practice: people prefer fairness to equality (Starmans et al., 2017).

7 Practices of Justice

7.1 Simplicity and Class Formation

Gigerenzer’s (2007, p. 199) statement is concise: “Simplicity is the ink with which effective moral systems are written.” He refers to the Ten Commandments as being incomplete, but clear, short and focusing on the essential. Similarly, the rules of Bentham and Kant are clear and simple and have lasted for centuries. Some very interesting observations regarding simplicity can be found in Schelling’s classic “The Strategy of Conflict” (1960). Focal points are outcomes where mutual expectations easily converge because they “enjoy prominence, uniqueness, simplicity, precedent […]” (p. 70). Schelling points to the “strong magnetism in mathematical simplicity” in international bargains and emphasizes the “remarkable frequency with which long negotiations […] converge ultimately on something as crudely simple as equal shares, shares proportionate to some common magnitude […] or the shares agreed on in some previous but logically irrelevant negotiation” (p. 67). This makes clear that the communicability of the
outcome is decisive. A focal point is a good reason for both sides, as are simplicity, symmetry etc. If some better reason can be found, then this will be the defendable, “reasonable” basis of distribution. In the debate about distributive justice, the principles of merit, need, and equality can be viewed as focal points, which might explain their prominent scholarly support in social psychology (cf. Section 1).

Another important way toward simplicity is class formation. Ideally, a just judgement would reflect the ‘true’ relation of costs and benefits of the case in question, but often neither is clearly determinable. A (mathematical) class is a group of elements that are not equal but can be viewed as equivalent regarding a certain aspect. Classes are abundant: pupils get grades, students make bachelor’s and master’s degrees (and get grades, too), boxers box in weight classes, worker’s payment is according to wage grades, VAT has certain tax rates, letters’ postage is according to weight classes and distance zones, parking meters charge full hours etc. The point here is that classes by definition are distinct to make unequal things comparable, be it for transparent pricing or equal chances or due assessment. Classes are used to make justice simpler and more efficient, but they have a few snares.

Take for example how cars are taxed. The bigger the car the higher the taxes, since bigger cars demand more public resources. But what does ‘size’ mean? Different countries have different approaches (cf. Wikipedia: “Road tax”), but they have in common that they define classes as distinct intervals of easily determinable values that can serve as proxies for size, e.g., engine displacement like 2001cc–2500cc, 2501cc–3000cc etc. as to have only a manageable number of classes and therefore a relatively simple system. Is it just? Well, it is contingent, but at least it is not unjust, practicable, communicable, and reliable.

Other important classes are nations (one of the most delicate class formations at all), for making very generalized reciprocity manageable, or the fiscal year, for finding a compromise between long-term fluctuations of profits (or income) and the need to come to terms with the past. The fiscal year is a practicable compromise, not the only possible, but surely a focal point, in contrast to nations where focal points (ethnicities) are often ignored and continue to be sources of endless struggle.

One of the most obvious problems of classes besides their general contingency are ‘absurdities’ along the class boundaries. It is easy to determine the engine displacement of a car, but difficult to reason whether a good is a ‘necessity’ and therefore to enjoy a lower VAT rate than a ‘luxury good’. Obvious cases are easy, like bread and jewelry, but what about sweets and clothes? Is it helpful to further differentiate between ‘necessary’ and ‘superfluous’ sweets and clothes? Tax regulations and their commentaries are infamous for their length and complexity, and some of them are absurd. But most of it is unavoidable, since generally the executive is bound to a strict procedural justice with transparent rules agreed upon in advance. These absurdities do not challenge the concept of class formation in general.

Justice is arduous, a never-ending balance between (brute) simplicity and (endless) differentiation.

### 7.2 Ideal Justice and Other Criteria

There are some places in Miller’s book (1999) where he casually, but confidently distinguishes justice and other criteria: “Even when genuine moral considerations – as opposed to simplicity or convenience – guide the choice of an allocation or a procedure, it may not be obvious whether justice, specifically, is what lies behind the respondent’s endorsements” (p. 47, original emphasis). Some lines further down he contrasts justice and social utility, somewhere else justice and efficiency (p. 216) or justice and consent (p. 104f.). These remarks are quite spectacular. He never defines what ‘justice’ is, and nobody else does so (at least not successfully) beyond “giving everyone his or her due”. Miller has an intuition of an ideal justice, and this intuition is generally reflected in the debate in terms like “genuinely moral”, “legitimate”, “adequate”, “justified”, or simply “fair”. Leventhal (1980, p. 52) illustratively contrasted “fair” and “quasi-fair” behavior, while Greenberg and Cohen (1982) would term it normative and instrumental justice. A famous example for this justice intuition is Radbruch’s formula, developed after World War II in reaction to German National-Socialist law practice (Radbruch, 1946). Radbruch objected the positivist identification of law and justice and emphasized the role of law ‘to serve justice’. But besides obvious violations – is pitting ideal justice against other criteria helpful in practice?

Viewed from the evolutionary perspective, long-term utility (functionality) is the decisive dimension of justice, but humans are not perfect. Yet because they are maximizers of utility, ostensibly utilitarian (functionalist, short-term) considerations can go awry. But the goal is the long-term participation in the social game. On the one hand, we must check ourselves via individual self-bindings: good intents, conscience, building up reputation, contracts, reliability etc. On the other hand, we must efficiently solve the free-riding problems of 1st and 2nd order: to recognize and sanction norm violators (Voland, 2013, p. 228). Practical justice must be efficient. It would not help to use up the resources that are to be defended by justice. Yet we must be careful to not to overshoot the mark while recognizing and sanctioning norm violators: “Shoot first, ask questions later” is a safe road to an unjust society. Therefore, a large part of our numerous self-bindings is not individual, but collective (institutional): fair trials, transparent administration, democratic decisions – procedures to avoid premature decisions biased to our personal advantage at the
expense of others and therefore maleficient in the long run.

To have an ideal of justice is necessary for analyzing ‘true’
merit or guilt and for demarcating it from additional criteria
and functionalist intentions that try to push efficiency or out-
comes (very popular: ‘jobs’) into the focus of the debate while
neglecting the basic problems. But in practice a principled dis-
tinction between ‘noble’ justice and ‘mundane’ considerations
is difficult. Ideal justice is the default here, but good reasons
can override the default.

8 The Implementation Problem

8.1 From Moral Will to Institutions

How is a Theory of Justice successfully implemented? Trad-
tional moral philosophy and common sense focus on indi-
vidual actions, assuming a duality of “genuinely moral mo-
tives” and “all other motives” (cf. Homann 2014, critically
discussing this discourse; cf. also Trivers 1971, p. 51). Moral
weakness is a weakness of moral will, the answer being “moral
rearmament”, i.e., developing better moral reasoning to resist
temptations.

Ulrich (2016, p. 451), for example, contrasts economy and
morality, and morality would require a breach of economic
logic. Homann (2003, 2014) is a critic of such a dualistic
opposition: Moral behavior must be advantageous for the
individual if he or she is not to be exploited by others who are
less moral and would initiate a moral downward spiral. Any
successful implementation of ethical rules therefore must be
compatible with economic incentives. Homann (2003, p. 226)
warned against regarding implementation “only” as a practical
problem: implementability precedes moral validity. An ethic
that presents unrealistic challenges to its participants is not
impractical but unethical.

We can add another dimension:

Enforceability: helpless ~ liberal ~ coercive

Justice must be reluctant, because only in liberty can people
develop to the full, deciding themselves what is best for them.
Self-appointed guardians of virtue may not restrict the freedom
of others, and paternalism or “nudging” are rightly criticized,
as is governmental overregulation. But justice cannot be tame.
The collective has a right to define justice and to enforce it.
Liberty without coercion (anarchistic self-control) is unrealis-
tic because people can drift far into self-righteousness (Haidt,
2013) and make their own rules – especially when rich or
“big”.

Brennan and Buchanan (1985) made a distinction between
choices within rules (choosing actions) and choices of rules
(designing constitution and laws). Homann (2014, p. 62)
termed these moves in the game and rules of the game. The
rules of the game are the institutional framework governing the
economic behavior in society – in short: the “economic order”,
a term coined by ordo-liberalism, a neo-liberal movement in
Germany in the mid-20th century (Eucken, 1992). Homann
(2003, p. 9) locates the systematic position of ethics in modern
(anonymous, pluralistic) societies in this economic order. The
rules must be moral as to enable the “players” to be guided in
their moves only by economic considerations, since economic
profits are a valuable indicator of sources of societal wealth
(Homann, 2014, p. 99). It is only this compatibility of sanc-
tioned rules and economic incentives that makes the system
inherently stable. The most important rule is to guarantee fair
competition in market economies (Eucken, 1992; Homann,
2003; Mankiw and Taylor, 2011). Individual ethics are never-
theless still important for filling the many gaps left out by the
economic order.

8.2 The Dark Side of Competition

The significance of fair competition for trade is widely ac-
nowledged, yet current political practice often looks differ-
ent. Governmental overregulation is only one part of “eco-
nomic policy”: attempts to influence or correct the outcomes
of “free” markets, with subsidies, protectionism, welfare pro-
grams, growth policies, comprehensive labor and employment
law etc. So, what is going wrong with fair competition?

Worldwide, economic growth is a primary goal of politics.
Richters and Siemoneit (2019a) recently argued that a political
growth imperative exists because politics would strive for jus-
tice, and that much of what is considered to be fair competition
is quite the opposite. They point to the fact that productivity
increases due to technological progress are not only the result
of “good ideas” but have an underestimated material basis
(Ayres and Warr, 2009; Kämmel, 2011; Ayres, Bergh, et al.,
2013). Accordingly, they distinguish a “genuine” form of com-
petition based on performance, when extraordinary personal
abilities and efforts (“merits”) are honored, and an innovation
competition, when not the innovation as such but mainly its
resource use leads to a better cost-benefit ratio of products
and services and, accordingly, to market success (Richters and
Siemoneit, 2019a, p. 134). Technology then undermines the
meritocratic principle by literally using resources not based
on merit. This systematic distortion of fair competition allows
technical occupations not only to bias the income distribution
forward their own benefit but is responsible for the continuous
threat of “technological unemployment”.

A political growth imperative in their view is created by the
interplay of three factors (p. 133): (1) increases in labor
productivity, leading to technological unemployment, (2) the
societal obligation to guarantee at least a minimal standard of
living for everyone (“need”), and (3) the meritocratic principle
as a fundamental social norm that sets limits to redistribution
between those earning a sufficient income and those who are
not.
8.3 Concentrating on “Non-Merit”

Section 2.2 described the wrangling over the meaning of merit. Participants in a study by Neckel et al. (2008, p. 45) found it easier to agree on what merit is not. This suggests using the exclusion of non-merit as an effective implementation of the meritocratic principle. From an Aristotelian perspective this would mean not struggling with the broad mean between extremes, but concentrating on the extremes themselves.

Richters and Siemoneit (2021) presented a possible solution to the problem of how to implement the meritocratic principle regarding the side conditions of effectiveness, efficiency, communicability, and liberty. In economics, undeserved incomes without a corresponding creation of wealth are called “economic rents”. Thus, justice in modern, democratic market societies would first and foremost mean to institutionally drain the wellsprings of such economic rents. They extensively discuss what in their view are two important sources of economic rents: (1) income shares which are substantially based on resource-intensive technologies, and (2) land rents, where a value generated by society is sold on private account. Appropriate institutions are caps on the extraction or import of non-renewable natural resources and a land value tax. Without economic rents, market economies could come closer to the goal of just self-regulation (Richters and Siemoneit, 2019b).

9 Discussion

9.1 Reconciling Altruism with Reciprocity

I have argued that the sociobiological notion of reciprocity can be regarded as the ideal of justice, and that an overwhelming, though often inconspicuous evidence indicates that reciprocity indeed governs all social relations – with the caveat that the term reciprocity seems to be an inappropriate framing for a substantial number of social situations (cf. Figure 3).

The more generalized and the less obvious reciprocity is, the more difficult it is to argue that certain actions are still part of an individual maximization of utility in the long run. But we must take seriously the sociobiological account that altruism in its selfless interpretation could not have been selected for in the evolutionary process. Although the term altruism may still have some justification in everyday language, for systematic reasons we should interpret altruistic acts as “blind investments” into society with a questionable “profitability” – hence, assuming reciprocity. With reference to Sahlins (1965) I have suggested to view such acts as “very generalized reciprocity”. We do not “need” altruism to explain stable patterns of just cooperation in societies.

In practice, justice is a social bargain and an optimization problem, with some of the effects depicted in Figure 4: Achievements and income serve as proxies for deserts and rewards, but our fellow citizens should not fall below the socio-cultural breadline. When individual efforts in groups are roughly comparable, their members may not argue about subtleties but prefer equality.

If we assume that the central goal is utility maximization, then balancing reciprocal expectations becomes only one goal among others that affect costs and benefits. Often maximizing net utility means minimizing (social, invisible) costs rather than maximizing (material, visible) benefits.

9.2 The Problems of Equity Theory

To approach a general theory of human interaction was indeed the claim of Equity Theory, but the way the theory was once presented (and defended) was probably infelicitous. Equity Theory would allow both signs (positive and negative) for inputs and outcomes, which made the concept confusing, and justified criticism was levelled where the meaning of inputs and outcomes had been repeatedly stretched only to achieve a balance, making the terms meaningless (e.g., Schwinger 1980, 98f., 105f., Deutsch 1985, p. 30, Reis 1986, p. 189). The mathematical formula presented conveyed the idea of precision and applicability but could never deliver on its promise (cf. Folger, 1986).

The basic problem I see is that the social notion of “inputs” and “outcomes” refers implicitly to the visible (communicable) preconditions and effects of social exchange. The notion of costs and benefits with their definite sign and “holistic” meaning is much clearer. There will never be a “quantification” of equity (just another word for full communicability), as demanded by Adams and Freedman (1976), when the terms of the underlying formula are difficult to quantify, strategically used, and stretched or “tugged” between conflicting interests. But there could be a theoretical agreement on its underlying principle.
9.3 Monism as Preferable Approach

Concerning the criticized “unidimensionality” of Equity Theory, I would regard it a virtue, not a failure, always keeping in mind that an operationalization of the meritocratic principle will only be possible by concentrating on non-merit. We need unidimensional theories to discover the relevant forces, i.e., a (strong) default and (weaker) deviations. That does not mean that social action gets uniform or “deterministic”, or that society would be “simple and free of contradictions” (Schwinger, 1980, p. 105, ironically). Rather we have seen that highly sophisticated strategies and arguments are used, often exactly for diverting the focus from the main dimension reciprocity to important other criteria. But basically, these remain to be “other” criteria.

9.4 The Relation of Distributive and Procedural Justice

It would be interesting to revisit the relation of distributive and procedural justice with the idea of reciprocity as the primary principle of justice. When outcomes are decisive, what is the systematic role of procedures? Social psychology regards both to be important dimensions without hierarchizing them (Lind, 2020). I would rather assume that insisting on certain procedures could be better explained as self-bindings for maintaining impartiality, avoiding biases on the part of the allocator. This would entail a primacy of outcomes, making procedures “only” part of the communication about justice and their outcomes.

9.5 The Social Construction of Merit

Reciprocity – especially regarding income for work – translates into the social norm Meritocratic Principle whose basic idea is to ensure rewards for your own achievement. Strictly speaking, there is nothing like “your own achievement”. The principle of the natural cycle is that all organisms appropriate calories “at the expense of others”, usually entailing the death of the organisms that have built them up. All creatures must choose their food (prey) according to whether its acquisition delivers more calories than it costs. To paraphrase an adage of Marxism, we all “make more calories from calories”, and the “profit” (excess calories) is used up for self-preservation and also (sociobiologically: first and foremost) for reproduction.

Therefore, Rawls and many others are not plainly wrong in their assumption that “merit” is a social construction. But it is not an arbitrary one. First, merit has a non-negotiable core, especially when time or quantities are the objective yardstick of achievement or when some people achieve results that others cannot, no matter how much effort they invest. Second, this construction has a deeper social meaning. Merit is the normative basis of the division of labor since nobody can achieve unboundedly (this is why we divide labor). Effort (i.e., hours spent on work) is an income-limiting factor, because only in cooperation with others can the talents of the few unfold and create specialists who deserve higher wealth. Their productivity is contingent on their social environment (team, firm, society, . . .) – hence, how is “high performance” to be attributed?

The meritocratic principle has a strong egalitarian element that is often not allowed for in the debate. There has been considerable resistance to the notion that the relationship between income and economic deserts in unregulated markets is just (e.g., the discussion in Olsaretti, 2004), and I fully agree. But Richters and Siemoneit (2021) have exhibited a way out of these troubles by showing how justice in markets can be achieved with few but strong and consistent measures. One of these are limits to inequality because inequality can develop a runaway dynamic far beyond any merit. From the collective perspective, endless accumulation is (and always has been) a social misunderstanding (though not from the individual perspective of course).

9.6 Focusing on the Essentials

A monist approach to justice based on biological facts and a common, basically uncontroversial practice has yet another advantage: it is a Theory of Justice that itself is normatively parsimonious. According to Gigerenzer (2007), effective moral systems should be clear, short and focus on the essentials, and this holds also for the institutions of modernity. The current philosophical debate often looks different (cf. Olsaretti, 2018): extreme positions try to avoid any bias (e.g., political correctness) and compensate for any perceived injustice (e.g., when couples prove unfertile and cannot have the children they wish). Instead, the societal quest for justice should concentrate on the blatant cases, i.e., “non-merit”, and should be content with results that are acceptable, i.e., “not unjust”.


