Rational Choice Modeling of Sustainable Corporate Social Responsibility

Werner Güth & Hartmut Kliemt

The intuition that corporate social responsibility CSR can serve as a functional equivalent of the moral conscience ascribed to personal actors can be restated in elementary Rational Choice Modeling terms. To serve this function, engaging inner constitutional commitments to act in the sense of CSR must be a viable option for corporate actors. Moreover, these commitments must be detectable by other actors with sufficient certainty at sufficiently low costs. Only where these requirements are fulfilled there is a niche for implementing the CSR ideals of stakeholder conceptions within the constraints of ordinary market competition and with respect for stakeholder property.

A. Introduction and overview

We share Christoph Engel's long-standing interest in 'trust'.[1] But like him and most others who approach trust problems within a rational choice perspective, we have mostly sidelined the fact that a kind of community of trustors and trustees who share certain aims, ends and values is typically involved. When trustors and trustees interact repeatedly then the growth or decline of trust of the trustors is not simply a matter of Bayesian updating of probabilities concerning the behavior of trustees.[2]

We can properly rely on the stability of a bridge but hardly properly trust in the bridge like we might trust in our physician as willing us well in using her discretionary powers as our agent even if there is no direct material incentive for this. But we cannot properly trust a company in the

1 Engel's views on 'generalized trust' in relation to experimental results see (Engel 2010; 2018); our own interest in the related general commitment problem arose from reading Selten (Güth) respectively Schelling (Kliemt). Our interest in modeling the evolution of trustworthiness resp. reliability started with (Güth and Kliemt 1994). We strangely enough completely missed out on the possibility of running an experiment on trust as in (Berg u. a. 1995)).

2 See on this in detail (Lahno 2002; 1995a; 1995b)).

same way as we trust a "significant other". Knowing this, we nevertheless speak and act in our relations with corporate actors as if we were continuously dealing with personal actors.

In this paper we allow for this way of speaking as well. Yet, we start from the tacit understanding that under the Rule of Law the reliability of corporations can be a functional equivalent of genuine trustworthiness. Where trustworthy natural persons are seen as intrinsically motivated by general moral rather than extrinsic material motives arising from the exigencies of separate choice-situations, corporations can be seen as being guided by internal constitutional rules. These rules are internal to the corporate actor and influence its behavior in external relations like commitments of personal actors.

We employ the 'lingua franca' of modern social theory or what we call "Rational Choice Modeling", RCM, which is mostly neutral with respect to the specific substantive Rational Choice Theory, RCT, employed, to illustrate how intrinsic can be represented besides extrinsic motivation explicitly in the same model of a corporate actor.[3] The ways and means of modeling intrinsic and extrinsic motivation in the case of personal actors within a "dual payoff function approach" that we have studied before can be transferred to the case of corporate actors. Thereby it becomes possible to provide a framework for studying the co-evolutionary process in which the population share of CSR actors develops in competition with non-moral profit-seeking personal and corporate actors. The upshot is that even in a world ultimately dominated by objective success – as measured in terms of profits and losses, spin offs and bankruptcies etc. – there can be a niche for decision-making that deviates from what measures of objective success would prima facie suggest. [4]

The indirect evolutionary approach models conditions under which CSR-oriented corporate actors can succeed in competition with purely profit-oriented actors even if we abstract of all personal relations by assuming random matching of interaction partners. This may be viable *provided* that internal corporate constitutional commitments to CSR can in principle be implemented and can be detected by potential interaction

3 For the relationship of RCM and, rational choice theory, RCT, see (Güth and Kliemt 2007).

4 To relate this to a rather popular management theory, we seek to represent an aspect of the general obliquity of (boundedly) rational human pursuits (Kay 2011) in precise RCM terms for the exemplary case of CSR.

partners. We believe that contrary to commonsense views firms as corporate actors have – at least potentially – a comparative advantage over personal actors when it comes to developing a functional equivalent of a conscience and to signal their 'firm commitments' transparently.[5]

CSR can *ultimately* be a case of doing well by doing good. To the extent that this is the case it renders CSR compatible with the *normative* demands of pursuing shareholder value in a private law society.[6]

B. Sustainable *CSR*

Like a virtuous personal actor who is by some intrinsic motivational process kept from following proximate extrinsic incentives, the good corporate citizen adopts corporate social responsibilities that seem directly unfavorable to the corporate actor's bottom line. As in the case of natural selection of virtuous individuals who are restricted by inner commitments in their opportunity taking behavior it seems initially puzzling how it can ever be (even in a wide sense of the term) "profitable" for a fully rational corporate actor to lack or to forego the ability to realize opportunities. After all, a fully rational actor endowed with "foresight and understanding" (Toulmin 1961) takes into account all future causal consequences of actions as foreseen by her in the instance of choice making. As a rational actor she knows that she will never choose a dominated alternative in the future. So, what might she gain by committing to do just that?

By assumption the rational actor does not suffer from weakness of the will and other imperfections. Therefore, she will be better off with more rather than fewer options no matter what. Since she will always be free to choose in view of *all* future consequences of her choices, she seems to have no use for commitments that restrict her future option sets. [7]

In a so-called game against nature avoiding myopic behavior would suffice. But it does not suffice in the case we are interested in. The selection

5 Alluding to the nice pun of (Mayer 2013).

6 Private law society in the sense of (Böhm 1989; 1966); a defense of an obligation to increase shareholder value as a demand implied by promises given within the legal order of private law societies can be found in (Kliemt 2022).

7 In his popular "economics in one lesson" Hazlitt describes the focus on all – including the most remote consequences of action – as the hallmark of rational economic behavior (Hazlitt 1988).

of behavior we are interested in, emerges from the behavior of other individuals who respond to restrictions of opportunism that they perceive in the actor.[8] This is the route we take to provide a positive answer to Robert Frank's challenge: "If Homo Ecomomicus Could Choose his Own Utility Function, Would He Want One with a Conscience?" (Frank 1987)?

I. An elementary example of how inner commitments may pay off

Imagine a simple exchange of apples and bananas (as in Figure 1). Person A has apples, *a,* and person B has bananas, *b*.[9] A prefers *b* to *a* while person B prefers *a* to *b*. Let the apples be ripe before the bananas. So, A needs to deliver his apples, *a,* first in the hope that B will deliver her bananas, *b,* afterwards:

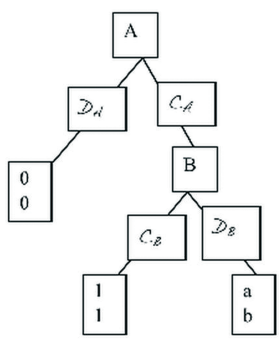

Figure 1 y>1>0>x represents players' preference rankings of plays of the game numerically
a, b, a&b, – in italics in the dotted boxes show substantial results of plays of the game

In Figure 1 a play of the game proceeds along a path from top to bottom. A chooses first, and B second. The numerical values at the end-nodes of the tree are *representations* of subjective preferences concerning the full path

8 See (Zahavi and Zahavi 1997) for instructive analogous examples from the animal kingdom.

9 One would have to think of a certain quality and quantity of apples and bananas at a certain time.

"all things considered". Of the pairs of rankings corresponding to plays of the game the upper entry shows the ranking by A and the lower the ranking by B.[10] The entries in the dotted boxes show as first entry of a pair what A possesses, as second entry the substantive holdings of B after the corresponding play of the game.

We have – left to right – at the origin of the tree two alternatives open to A: *left, D_A, or non-delivery of apples by A* whose ranking of this alternative is represented by 'utility', "0" (representing the subjective preferences of A); right , C_A, *or delivery of apples a by the first mover A*. After C_A the second mover B can deliver *b* by choosing move C_B – leading to a "*subjective*" overall "payoff" or utility of "1" for, B, and of "1" for A. Alternatively, the second mover B can choose D_B leading to the play (C_A, D_B) and a "subjective" overall payoff of "*y(>1)*" for B – holding a & b after the play (C_A, D_B) – and of "*x(<0)*" for A – holding nothing after the play (C_A, D_B).

If the 'all-things-considered' subjective ordering u_i, $i = A, B$ of the substantive results is the same as the 'natural' ordering ξ_i, $i = A, B$ that considers only substantive results, then a rational actor A who foresees B's evaluation *y>1* would never deliver his apples in the first place. Due to 'backward induction', he would foresee ending up with nothing at all after choosing C_A and therefore would choose D_A to begin with.

Still, both could be better off in naturally ordered holdings than in the status quo. This would be the case if the actor in the second mover role were able to

(i) commit to behavior that goes against the extrinsic motives exerted by the substantive payoffs in the dotted boxes and
(ii) could signal this commitment. If for instance, due to a commitment, the all-things-considered preferences represented by u_B *would* change to 'new' all-things-considered preferences which *could* be represented by u_B^* with $u_B^*(C_A, D_B) < u_B^*(C_A, C_B)$ then this would indicate that actor B –

due to the additional commitment – would prefer the path (C_A, C_B) over (C_A, D_B) even though $\xi_B(C_A, D_B) > \xi_B(C_A, C_B)$.

10 The scale factor and zero points of the numerical preference-representations have been chosen according to some order-preserving positive linear transformation that yields (0, 0) for the status quo and (1, 1) to the status reached by a successful exchange.

Once B can commit and signal her commitment and A can detect the presence of suitable intrinsic motives, backward induction will induce A to anticipate the co-operative play (C_A, C_B) as resulting from his choice. This result will be better than the status quo according to the evaluations of both actors.

So, if homo economicus could in a situation like that of Figure 1 choose his own utility function (s)he would choose one with a conscience.

This seems to answer the "Frank-question". Yet, though interesting it does not show that individuals endowed with a conscience would survive in evolutionary competition. In particular, the familiar problem of mimicry would have to be overcome since it could be better to appear as if having a conscience without having it 'in truth'.

We do not think that it is convincing to think in stark alternatives of non-detectable commitments and fully detectable commitments here. So, we will first make a few remarks about the range of parameter values for which a positive share of conscientious actors might conceivably be evolutionarily stable. The arguments concerning the sustainability of a positive share of committed actors seem precarious since in case of personal actors the mimicry problem is lurking in the background. However, as far as this problem is concerned CSR as a functional equivalent of a conscience seems different and less problematic because its presence may be observable. So, setting aside the problem of mimicry let us first turn to a brief sketch of the basic comparative statics of the adaptive process in which the disposition to "do good" can be sustainable. After that we shall return to mimicry and the sustainability of CSR.

II. A graphical representation of pre-requisites of CSR sustainability

When called upon to act in the second-mover or B-role in exchange relationships transaction partners can be either 'trustworthy/reliable' (characterized by $u_B(C_A, D_B) < u_B(C_A, C_B)$) or 'untrustworthy/non-reliable' (characterized by $u_B(C_A, D_B) > u_B(C_A, C_B)$). When called upon to play in the first-mover or A-Role a rational actor will decide on C_A iff $p_A u_A(C_A, C_B) > u_A(C_A, D_B)(1 - p_A)$;

where p_A is the subjective probability estimate of the player in the A-role that the partner in the B-role is trustworthy. If we assume (i) that p, $0 \leq p \leq 1$, is the share of trustworthy types in the population and (ii) that there is no way to tell the trustworthy ex ante from the non-trustwor-

thy then under purely random matching we have

$$pu_A\,(C_A, C_B) > u_A\,(C_A, D_B)\,(1-p)\,;$$

as a sufficient condition of offering cooperation in the trustor role. Yet, if information concerning B can be acquired after random matching by actors in the A-role at uniform material cost C, $0 \leq C \leq 1$, then this information will influence the rational expectations of players.

In Figure 2 one finds for alternative levels of costs C* horizontal arrows that indicate how p will evolve at that cost level from any starting combination (p*, C*).

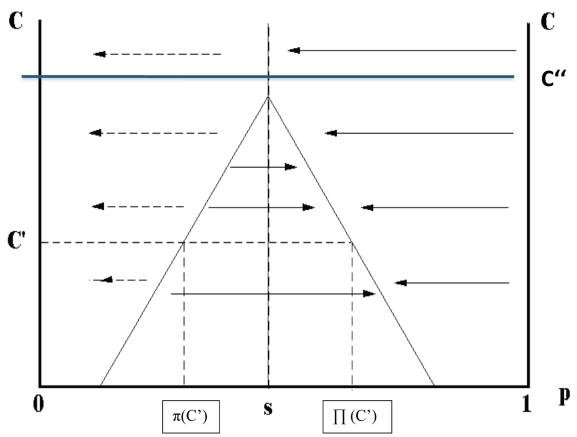

Figure 2 Evolution of population share p of committed actors for alternative cost levels C and given reliability of type detection

In Figure 2 values of p are indicated on the abscissa. On the ordinate we find alternative cost levels C at which a detection [signaling] [11] technology of a certain reliability is available to the interacting parties. Horizontal lines in the triangle indicate for alternative values of C the population

11 The detection and signaling technologies relevant here amount basically to the same thing. Either way a stochastic signal of the probability that the co-player is reliable respectively non-reliable is received by the trustor who can update her beliefs accordingly; (see for analytical details (Güth and Kliemt 2000)).

compositions for which it is worthwhile to employ a detection technology of given reliability; e.g. for C=C' the relevant interval reaches from $\pi(C')$ to $\sqcap(C')$.

The arrows indicate the direction in which p will evolve whenever evolutionary success monotonically depends on direct substantive success (proximate fitness) on each round of interaction. For all initial (p*, C*) combinations it can be characterized where the process will ultimately be heading for C*.

So, for C′ in Figure 2

if $p \in (\pi(C'), \sqcap(C'))$ the population share of the trustworthy will increase up to $\sqcap(C')$;
if $p < \pi(C')$, the population share of the trustworthy will decrease down to 0;
if $p > \sqcap(C')$, the population share of the trustworthy will decrease down to $\sqcap(C')$.

Generally, the higher the costs of the detection technology, the smaller the interval of p-values for which it is rational to invest in the costly technology. Only for generic intervals the given technology will render a positive share of committed actors sustainable.[12] Assuming an infinite population and random pairing the process will converge either, for $p > \pi(C)$ towards $p = \sqcap(C)$ and otherwise towards $p = 0$.

Due to the dual nature of the conceptual framework underlying the indirect evolutionary approach (i) the choices are made according to subjective expected utility u_A, u_B, yet (ii) the population share p_{t+1} is determined solely according to the substantive success ξ_A, ξ_B of these choices.

The parameter s denotes the population share p=s at which, without access to type information, the expected values of showing trust or not would be equal. Knowing only the population share p, yet not the type of the specific partner, *all* A-role actors will show trust if $p > s$ and will show no trust if $p < s$. Without specific type information untrustworthy types in the second mover role will do better than trustworthy types. The untrustworthy will crowd out the trustworthy if beyond knowledge of the population share p of trustworthy types, no type-information is available

12 Figure 2 shows the relevant triangle for a technology of given reliability. This technology makes two signals available, one of trustworthiness and one of non-trustworthiness. The signals may differ in reliability yet since nothing substantial hinges on this for the present purposes we leave it out of account.

at sufficiently low costs: As long as $p>s$, untrustworthy B-role actors will exploit the rational bets placed by A-role actors. If $p<s$, untrustworthy B-role actors will be offered a chance to exploit first-mover trust only if A-role actors erroneously show trust. Since the chances to exploit "noise trust", $p<s$, are less frequent than under conditions in which trust is the better choice for all A-role actors, the noise-dynamic is slower – indicated by dashed arrows – than the choice-dynamic – indicated by bold arrows. Obviously if A-role actors can make use only of their knowledge of the population composition p then ultimately the share p will converge towards $p=0$.

In sum, from our modeling approach we know that for any $C<C''$ not only $p=0$ but also $p>0$ may be an evolutionarily stable rest point. (For $C>C''$ the process converges to p=0 irrespective of the (initial) type composition.) A precise RCM makes it plausible that there can be a "niche" for intrinsically motivated moral types even in large scale competitive interactions.

III. Intermediate conclusion

According to the preceding line of argument the a priori thesis that market competition of necessity will drive out CSR investments is untenable.[13] Our example of pursuing an indirect evolutionary approach shows that it is at least conceivable that individuals who due to their commitments are merely boundedly opportunistic will not necessarily be crowded out by uncommitted competitors. Provided that individuals command the ability to detect committed actor-types with sufficient reliability at sufficiently low costs, commitments that express themselves in proximately disadvantageous CSR choices may indirectly contribute to the ultimate success of committed types.[14]

We are in principle content to let it rest with the insight that commitments to CSR can be sustainable under certain conditions.[15] Commit-

13 A parallel argument is made in (Güth and Kirchkamp 2021).

14 Dawkins observation that in case of personal actors long term survival is the greatest ultimate success carries over to the case of corporate actors as well; see (Dawkins 1995)

15 Whether the argument will work beyond mutually advantageous exchange and contract is open. But even if the viability of "doing well by doing good" were restricted to

ments to execute (incomplete) contracts not only according to the letter but in spirit – i.e. genuine trustworthiness rather than reliability – will be sought after in corporate actors as potential transaction partners as it is sought in personal partners. Yet, since "hunger is not bread"[16], the fact that commitments would be desirable does not prove their feasibility. We believe that contrary to common sense corporate actors may be seen as commanding a comparative advantage over ordinary personal actors when it comes to having a conscience or commitments functionally equivalent to it.

C. Corporate constitutions as "corporate conscience"

The dual function model underlying the indirect evolutionary approach remains externalist in focusing on overt behavior rather than on internal processes that generate behavior.[17] We now illustrate by concrete examples how empirically the RCM framework may be specified.

I. Creating and signaling general corporate constitutional commitments

Trust in an agent's (trustee's) intrinsic motivation to pursue the aims ends or values of the trustor (principal) can be substituted by explicit contract. But if complete contracts specifying substantive outcomes are either impossible or undesirable, potential interaction partners may place value on internal *procedures* of a corporate actor that provide a functional equivalent of a personal actor's trustworthiness (e.g. in using her discretionary powers according to Rule of Law principles). A simple example can illustrate how this may be accomplished in principle (though real practices will be more complex).[18]

---

exchanges and contracts that are directly advantageous to all involved it would still be a highly relevant result.

16 See ((Bentham 1843) fallacy II).

17 In the standard approach there is no room for intrinsic motivation per se. Since only the choice behavior resulting from motivational processes matters is suffices to represent the choices made – or so the standard argument runs.

18 The existence of procedural rules in practice, of course, requires fundamental rule following behavior of individuals. On experiments concerning the empirical side of rule following see (Kimbrough and Vostroknutov 2016; 2018; Gächter u. a. 2025).

II. Simple majority rule as commitment device[19]

Imagine a group of five individuals operating as a corporate actor under simple majority rule in some realm of choice. One individual is of age up to 1, one up to 2, one up to 3, one up to 4, and one up to 5. Once an actor becomes member he will be able to go for at least five periods. At the beginning of each period the oldest actor presently on the team, upon turning 5, exits from the community and is substituted by an individual of age 0. There is no exit other than the regular one after five periods and no new member is admitted unless a regular exit has taken place.

In our stylized example of procedural corporate governance, the age composition and the simple majority rule were fixed. The median voter is always aged less than three and has to go on for at least three periods of corporate choice making (including the one in which the decision is made). These aspects of the internal structure of the corporate actor may be fixed in externally enforceable contracts. But there is and must be residual discretionary power in collective decision making of the corporate actor. All individuals who have a voice in the decisions retain their powers of judgment and the freedom to vote for or against proposals, respectively, as seems fit to them. *Decision procedures rather than decisions are fixed.* The procedure sees to it that when exercising their discretionary power, the median and all younger voters must always consider at least two future periods of interaction. Therefore, the collective decisions of the corporate actor are not suffering from the same "end-game" effects as decisions of individual actors. Incentives to show opportunity taking behavior on the corporate level are in this sense bounded even if the individuals act fully opportunistically.

Individuals who join such a corporate actor know that the group as a whole will be committed at every point in time to adopt a perspective that goes beyond the short term. Personal or corporate actors who face such a corporate actor in external relations also understand that the actor is substantively unrestrained but will use its discretionary power in a way that reaches beyond the direct consequences for each single round of play taken separately even if the personal actors "within" the corporate actor are acting as opportunistically as in the outside world.

19 For the simple modeling idea originally due to Brennan see (Brennan and Kliemt 1994), for more general and sophisticated treatments of the folk theorem logic involved, see e.g. (Kandori 1992), (Salant 1991).

In sum, we have both, discretionary powers and a restriction on the nature of the opportunity set that will induce less short-term oriented and in that operational sense more "responsible" behavior.

Artificial and simple as the example may be, it nevertheless illustrates how procedures of collective decision-making "within" a corporate actor can serve as a commitment device. Of course, the workings of the mechanism depend on being committed to playing by the rules. To the extent that people have reasons to believe in rule of law and can rely on the legal order they can also have good reason to rely on the external enforcement of procedural rules internal to the corporate actor.[20] They can *rely* on the fact that the collective actor will have the internal age structure described before, decides by simple majority, and has a potentially infinite life if it can sustain itself. Thereby the externally monitored internal corporate governance structure can induce external interaction partners of the corporation to expect corporate social responsibility in the sense of its bounded rather than unrestricted opportunism. This is a clear sense in which a functional equivalent of trustworthiness may be created by an appropriate constitutional decision architecture.

Internal commitments to certain procedures of corporate governance are not only relevant for external relations of the corporate actor. They are also relevant for its relations to its own members. Again, an extremely stylized simple example may serve as an illustration of the potential advantages of such forms of *procedurally embodied* corporate social responsibility.

III. Unanimous co-determination of decisions

Other than shareholder, so-called stakeholder approaches argue that managers and the companies they run should be treating all potential interaction partners of the company respectfully and fairly in an ethically non-discriminatory manner.[21] Whether this can plausibly hold for external

20 We admit that institutionalized constitutional commitments to "the political ideal of the rule of law" (Hayek 1955) ultimately depend on *informal practices* of individual rule following. See on the nature and significance of the political ideal of the rule of law (Kliemt 2024).

21 On Freeman's influential variant of such an approach see (Freeman 1984), (Freeman u. a. 2010).

relations with suppliers and customers as stakeholder conceptions often insinuate seems doubtful.[22] With respect to internal relations with employees or shareholders such an ideal of non-discriminatory treatment may seem more plausible and can indeed be spelled out in terms of unanimous decision-making.[23]

The unanimity rule must in any event seem very attractive to adherents of stakeholder conceptions of corporate social responsibility since it implies that shareholder interests do not command "ethical" priority vis a vis stakeholder-interests. It expresses the stakeholder conceptions' core ethical value of equal moral concern (operationalized as equal veto power) for all stakeholder interests in concrete institutional rule terms.[24]

Though it is natural for an ethicist to subscribe to norms of equal ethical respect this subscription rests uneasily with the privileges of private property and ownership in WEIRDS (Western, Educated, Industrialized, Rich, Democratic Societies) which in all likelihood are part of the secret of the success of such societies (Henrich 2020; 2016). It also disregards the fact that contracts with other individuals on behalf of the corporation in the last resort derive their authority from being authorized by residual claimants or owners. If we are not willing to brush these rather obvious *normative* objections aside then following the precepts of a stakeholder approach, to the extent that it can be justified at all, must be justified indirectly: to adopt a stakeholder perspective in 'human resource management' must pay off over the long haul for shareholders of the corporation. Unless it is a variant of 'doing well by doing good' it is hard to see how it could be compatible with respecting property rights of shareholders under the rule of law.[25]

Despite this, as far as relations to employees are concerned a corporate culture in which the values of fairness and equal respect for each and

22 See on some of such doubts in a management context (Kliemt 2022)

23 Of course, the complementarity of formal and informal "rules" in corporate culture is acknowledged; see for instance succinctly on this the comments in (Engel and Güth 2006)

24 See on the ways and means of spelling this idea out (Brennan and Kliemt 2018; 2022; Kliemt 2021).

25 The exit option plays a crucial role here: if management uses the discretionary powers vested in it in ways that keep the share value up, those shareholders who disagree with its decisions get the best start when selling and using their own resources for their own aims outside the corporation. Likewise, employees can exit the employment nexus of a corporation; see for related issues (Buenstorf u. a. 2016).

everybody are guiding principles of interaction may seem to have great promise. No wonder that senior management expresses support for such a corporate culture in speeches. However, potential employees may demand something more reliable than speeches in which principles of corporate social responsibility in dealing with stakeholder employees are endorsed merely as lofty ideals. To win (potential) employees over, the implementation of corporate governance procedures that *embody* equal respect may be necessary.[26] But as always, the devil is in the details, and we have a hunch that the devil might ultimately have the upper hand here.[27]

D. Final observations

To amount to anything real, "stakeholder-talk" must be translated into "stakeholder rules"[28]. Majority rule though giving an equal vote to everyone does not express equal respect sufficiently in procedural terms. As has been observed through the ages majorities may overrule minorities the same way as bosses in a hierarchy overrule their underlings. Taking stakeholder conceptions at their word only the strict unanimity rule could express guarantees of equal concern procedurally and in institutional terms.[29]

Nobody would suggest in earnest that all decisions in real world affairs should be made unanimously by all affected individuals. Only for certain fundamental decisions veto power may be procedurally granted to individuals of a certain group of stakeholders (or to representatives of homogenous stakeholder groups). Though in some circles of contractarian political philosophers and contractarian political economists, *conceivable*

26 Acquiring firm specific human capital can be conceived as analogous to showing trust in the interaction of Figure 1. Putting the employer in a role analogous to the B-role actor of Figure 1 the A-role employee has to be aware that his exit option may decrease in value and the employer's bargaining position may be strengthened (in particular the threat point of negotiations may alter in the course of time in ways unfavorable to the specializing employee).

27 See for an assault on the problem (Güth 2014).

28 Implementing deferred bonuses would be another obvious way to credibly signal a commitment of the corporation and its management to the longer-term

29 Though protection by substantive contractual claims and rights may be useful they can only regulate what is anticipated.

unanimity is regarded as a useful tool of conceptual analysis a merely conceivable is not a real-world procedure (hunger is not bread).[30]

Due to the elegance of the revealed preference approach economists became almost oblivious of the fact that stenographical representations of the results of decision-making will work well only for understanding contexts in which the internal springs of action do not matter much.[31] This will be the case if observed behavior is – as in case of some markets – mostly determined by the *structural* conditions of *inter*action and the substantive or material payoffs that characterize them (typically in monetary terms). [32] In such cases of market competition actors who decide against what substantive payoffs require will be swiftly eliminated and economists can feel safe in ignoring behavior that is not substantively advantageous (one of the reasons why we modeled the evolutionary niche for CSR in so much detail).

It should be noted well, though, that the unitary actor model will suffice only under the special conditions described before. Only then can the psychology be taken out of economics without losing much. But if we consider more complex forms of behavior that at least allow for trade-offs between the short and the long term we cannot avoid shedding some light onto how behavior is generated.

Looking at internal processes of corporate actors we can treat their internal decision processes as equivalent to mental processes of personal actors who are cognitively and emotionally bounded (committed) to some

30 Even the most ardent adherents of unanimity try to cope with this by making its operation more indirect, for the seminal analysis see (Buchanan and Tullock 1962) and the still devastating criticism of classical contractarianism in "of the original contract" (Hume 1985).

31 Some of the most interesting work in making economic models of individual behavior more adeaquate make use of such techniques. To name but two examples one might think of (Bolton and Ockenfels 2000) as well as (Fehr and Schmidt 1999) which both build distributional and other regarding concerns into "individual utility function". We have no objection against devoting attention to motivational processes internal to the actor. We resent, however, insisting that human behavior must by all means be described as if maximizing some function. This can be justified only if one has – as in a context of evolutionary competition – good reason to assume that inconsistent choice making that cannot be so represented will be selected against strongly.

32 It is not by chance that the models in which economists tend to take the greatest pride "predict" certain outcomes of interaction mostly independently of specific assumptions on the behavioral level; e.g. evolutionary models as again in (Schumpeter 1959), (Alchian 1950), experimental markets (Smith 1962), or chess board models like in (Schelling 1971), (Hegselmann 2012).

extent. Both, corporate and personal actors need not be utilitarian consequentialists. They might as well start from premises of interpersonal respect of a broadly speaking Kantian type. Whether such views are ethically reasonable or not is discussed in normative ethics. But it seems beyond reasonable debate that Corporate Social Responsibility can be a serious corporate governance strategy only if it can be institutionalized in operational terms.

First, it must be shown how CSR can be embodied in institutional rules of corporate governance.

Second, institutionalized CSR commitments must be shown to be sustainable.

To the extent that we can by means of corporate governance prepare at all for future decision situations that cannot yet be anticipated it should not be in terms of prescriptions of substantive choices.[33] Neither personal nor corporate actors will want to be operating according to a fully specified explicit strategy leaving no discretionary power to future manifestations of the actor. Commitments therefore must concern the procedures rather than the substance of choices. Why CSR should be an exception to this is as hard to see as why it should be impervious to the strictures of sustainability in competition. On the other hand, RCM can be used to make it entirely clear what it takes that CSR can be sustainable in market competition and how the implicit managerial promise to hold up share values can be compatible with CSR. In defense of shareholder value conceptions we should not forget that shareholders are also stakeholders. Strengthening the shareholders' exit option by keeping the share value up is expressive of *ethical* respect for the plural values of this particular class of stakeholders who can always invest their capital in pursuit of their particular aims, ends and values with some competitor.

33 The enumeration of several mutually exclusive sufficient action triggers can always be made collectively exhaustive by an "else-clause" specifying the corresponding action. But no person in her right mind would believe in her ability to anticipate "everything else" sufficiently to warrant a substantively fixed response, *no matter what*.

References

Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory". *Journal of Political Economy* Vol. 58: 211–21.

Bentham, Jeremy. 1843. *Anarchical Fallacies*. Herausgegeben von John Bowring. Bd. 2. The Works of Jeremy Bentham. William Tait. https://oll.libertyfund.org/titles/bentham-works-of-jeremy-bentham-11-vols.

Böhm, Franz. 1966. "Privatrechtsgesellschaft und Marktwirtschaft". *ORDO* 17: 75–151.

Böhm, Franz. 1989. "Rule of Law in a Market Economy". In *Germany's Social Market Economy: Origins and Evolution*, herausgegeben von Alan Peacock and Hans Willgerodt. Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20145-7_4.

Bolton, Gary, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity and Competition". *American Economic Review* 90: 166–93.

Brennan, Geoffrey, and Hartmut Kliemt. 1994. "Finite Lives and Social Institutions". *Kyklos* 47 (4): 551–71.

Brennan, Geoffrey, and Hartmut Kliemt. 2018. "Kantianism and Political Institutions". In *The Oxford Handbook of Public Choice, Volume 1 (Oxford Handbooks) (English Edition): Roger D. Congleton, Bernard N. Grofman, Stefan Voigt (eds.)*. Oxford Handbooks.

Brennan, Geoffrey, and Hartmut Kliemt. 2022. "Politics as Exchange?" *Homo Oeconomicus*, Online-Vorab-Publikation, August 17. https://doi.org/10.1007/s41412-022-00128-5.

Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent*. University of Michigan Press.

Buenstorf, Guido, Christoph Engel, Sven Fischer, and Werner Gueth. 2016. "Non-compete clauses, employee effort and spin-off entrepreneurship: A laboratory experiment". *Research Policy* 45 (10): 2113–24. https://doi.org/10.1016/j.respol.2016.08.005.

Dawkins, Richard. 1995. *River out of Eden. A Darwinian View of Life*. Science Masters. Weidenfeld & Nicolson.

Engel, Christoph. 2010. "The behaviour of corporate actors: How much can we learn from the experimental literature?" *Journal of Institutional Economics* 6 (4): 445–75. Cambridge Core. https://doi.org/10.1017/S1744137410000135.

Engel, Christoph. 2018. "The Proper Scope of Behavioral Law and Economics (January 1, 2018)." MPI Collective Goods Discussion Paper No. 2018/2,. Bonn. http://dx.doi.org/10.2139/ssrn.3104475.

Engel, Christoph, and Werner Güth. 2006. "Corporate Culture as a Resource for Management: Comment". *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 162 (1): 97–100. JSTOR.

Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics* 114: 817–68.

Frank, Robert. 1987. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *The American Economic Review* 77/4: 593–604.

Freeman, R. Edward. 1984. *Strategic Management: A Stakeholder Approach*. Pitman.

Freeman, R. Edward, Jeffrey S. Harrison, Andrew C. Wicks, Bidhan L. Parmar, and Simone de Colle. 2010. *Stakeholder Theory: The State of the Art*. Cambridge University Press.

Gächter, Simon, Lucas Molleman, and Daniele Nosenzo. 2025. "Why people follow rules". *Nature Human Behaviour*, Online-Vorab-Publikation, Mai 26. https://doi.org/10.1038/s41562-025-02196-4.

Güth, Werner. 2014. "Endogenous community formation and collective provision – A procedurally fair mechanism". *Journal of Economic Behavior & Organization* 108 (Dezember): 389–95. https://doi.org/10.1016/j.jebo.2014.01.005.

Güth, Werner, and Oliver Kirchkamp. 2021. "Believing in Corporate Social Responsibility: An Indirect Evolutionary Analysis". *Journal of Institutional and Theoretical Economics*.

Güth, Werner, and Hartmut Kliemt. 1994. "Competition or Co-Operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes". *Metroeconomica* 45 (2): 155–87. https://doi.org/10.1111/j.1467-999X.1994.tb00018.x.

Güth, Werner, and Hartmut Kliemt. 2000. "Evolutionarily Stable Co-operative Commitments". *Theory and Decision* 49 (3): 197–222.

Güth, Werner, and Hartmut Kliemt. 2007. "The Rationality of Rational Fools.The Role of Commitments, Persons and Agents in Rational Choice Modeling". In *Rationality and Commitment*, herausgegeben von Fabienne Peter and Hans Bernhard Schmid. Oxford University Press.

Hayek, F.A. 1955. *The Political Ideal of the Rule of Law*. Fiftieth anniversary commemoration lectures. National bank of Egypt. https://books.google.de/books?id=5REeAAAAMAAJ.

Hazlitt, Henry. 1988. *Economics in One Lesson: The Shortest and Surest Way to Understand Basic Economics*. Crown Business.

Hegselmann, Rainer. 2012. "Thomas C. Schelling and the Computer: Some Notes on Schelling's Essay 'On Letting a Computer Help with the Work'". *Journal of Artificial Societies and Social Simulation* 15 (4): 9.

Henrich, Joseph. 2016. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Henrich, Joseph. 2020. *The Weirdest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. 01 Edition. Allen Lane.

Hume, David. 1985. *Essays. Moral, Political and Literary*. Liberty Fund.

Kandori, Michihiro. 1992. "Repeated games played by overlapping generations of players". *The Review of Economic Studies* 59 (1): 81–92.

Kay, John. 2011. *Obliquity. Why our Goals are best achieved indirectly*. Profile Books.

Kimbrough, Erik O., and Alexander Vostroknutov. 2016. "Norms Make Preferences Social". *European Journal of Political Economy* 14 (3): 608–38.

Kimbrough, Erik O., and Alexander Vostroknutov. 2018. "A Portable Method of Eliciting Respect for Social Norms". *Economics Letters* 168: 147–50.

Kliemt, Hartmut. 2021. "The logical foundations of constitutional democracy between legal positivism and natural law theory". *Public Choice* 195 (3): 269–81. https://doi.org/10.1007/s11127-021-00888-9.

Kliemt, Hartmut. 2022. "What should moral managers do? The moral virtues of shareholder value conceptions of management". *The'Betrieb'as Corporate Actor*, 108–21.

Kliemt, Hartmut. 2024. "The nature and sicgnificance of the Rule of Law". *Kyklos* 77 (4): 1084–102. https://doi-org.ezproxy.uni-giessen.de/10.1111/kykl.12407.

Lahno, Bernd. 1995a. "Trust, Reputation, and Exit in Exchange Relationships". *Journal of Conflict Resolution* 39 (3): 495–510.

Lahno, Bernd. 1995b. *Versprechen: Überlegungen zu einer künstlichen Tugend*. Oldenbourg Wissenschaftsverlag.

Lahno, Bernd. 2002. *Der Begriff des Vertrauens*. Mentis.

Mayer, Colin. 2013. *Firm Commitment: Why the corporation is failing us and how to restore trust in it*. Oxford University Press.

Salant, David J. 1991. "A repeated game with finitely lived overlapping generations of players". *Games and Economic Behavior* 3 (2): 244–59. https://doi.org/10.1016/0899-8256(91)90025-A.

Schelling, Thomas C. 1971. "Dynamic models of segregation". *The Journal of Mathematical Sociology* 1 (2): 143–86. https://doi.org/10.1080/0022250X.1971.9989794.

Schumpeter, Joseph Alois. 1959. *The Theory of Economic Development*. Harvard University Press.

Smith, Vernon L. 1962. "An Experimental Study of Competitive Market Behavior". *Journal of Political Economy* 70:2 (April): 111–37.

Toulmin, St. 1961. *Foresight and understanding. An enquiry into the aims of science*. Hutchison.

Zahavi, Amotz, and Avishag Zahavi. 1997. *The Handicap Principle. A Missing Piece of Darwin's Puzzle*. Oxford University Press.0.