Hartmut Kliemt*

On the Nature and Significance of (Ideal) Rational Choice Theory

https://doi.org/10.1515/auk-2018-0006

Abstract: The increasingly wide spread use of RCM, rational choice modeling, and RCT, rational choice theory, in disciplines like economics, law, ethics, psychology, sociology, political science, management facilitates interdisciplinary exchange. This is a great achievement. Yet it nurtures the hope that a unified account of rational (inter-)active choice making might arise from 'reason' in (a priori) terms of intuitively appealing axioms. Such 'rationalist' characterizations of rational choice neglect real human practices and empirical accounts of those practices. This is theoretically misleading and practically dangerous. Searching for a wide reflective equilibrium, WRE, on RCT in evidence-oriented ways can explicate 'rational' without rationalism.

Keywords: rationalism and unity of science, rational choice, bounded rationality, critical rationalism, Goodman's induction, Selten's methodological dualism, Weber's ideal types

1 Introduction and Overview

Emergent from common Hobbesian roots of theorizing 'more geometrico' about how people would and/or should behave modern (axiomatic) rational choice theory—RCT, henceforth—and its formal *language* of rational choice modeling—RCM, henceforth—exist since mid 20th century.[1] The increasingly wide spread use of RCM and RCT—in disciplines like economics, law, ethics, psychology, sociology, political science, management—facilitates interdisciplinary exchange.[2]

1 I am aware, of course, that RCT also stands for randomized controlled trials which form one of the pillars of proper evidence oriented research on 'moral', 'medical', 'political', 'managerial' etc. 'subjects'. The use of RCT in the rational choice sense is orthogonal to randomized controlled trials.

2 RCM has the invaluable merit of forcing researchers to explicitly formulate their assumptions about what is and what is not subject to the causal influence of choices within the rules of a given

***Corresponding author: Hartmut Kliemt,** c/o Prof. Dr. Max Albert, Volkswirtschaftslehre, Justus-Liebig-Universität Gießen, e-mail: hartmut.kliemt@t-online.de

This is a great achievement. It nurtures the hope that a unified account of rational (inter-)active choice making might arise from 'reason' in (a priori) terms of intuitively appealing axioms. 'Rationalist' characterizations of rational choice that neglect real human practices and empirical accounts of those practices are, however, theoretically misleading and practically dangerous.[3]

To illustrate the achievements, limits and risks of RCT the next two sections will locate both RCT and my discussion of it on a coarse intellectual map of science and philosophy (*section 2 and 3*). This sets the stage for a critical assessment of the scope and limits of Reinhard Selten's methodological dualism and his strict separation of 'ideal' (a priori) and 'real' (a posteriori) RCT (*section 4*). The next section opts for the unity of science by interpreting RCT as a nomological discipline focusing on the explanation of 'rational' behavior in empirical terms other than its rationality (*section 5*). Then I try to cope in critical rationalist terms with the notorious question of how the descriptive interact with the prescriptive dimensions in our search for a wide reflective equilibrium, WRE (*section 6*). Summary conclusions end the paper (*section 7*).

2 Putting RCT into Perspective

The graphical overview of this section (graph 1) and the comments concerning it are formed according to the kiss—keep it simple stupid—principle. Without claiming to present a fully-fledged philosophical argument they illustrate possible relations between 'mother philosophy' and 'moral science' as its (il-)legitimate social science offspring was once called. Their sole purpose is that of providing a coarse account of where—within a history of ideas context—the subsequent discussion is located on the 'intellectual map'.

I distinguish between philosophical approaches to human behavior that follow the philosophical tradition of using the terms 'philosophy' and 'science' mostly synonymous and approaches that strictly separate philosophy and science. Endorsing a broadly Humean view that emphasizes the continuity between

interaction (game) and the law-like hypotheses by which consequences of action are predicted; see in detail on RCM vs. RCT, Güth/Kliemt 2007.

3 The entertaining but often absurd 'story telling' in so-called economic imperialism forms a relatively harmless case in point. At best it can deliver interesting explananda but no empirical explanations; see McKenzie and Tullock 1978. Equilibrium assumptions like that of 'arbitrage free financial markets' will in combination with the apparatus of ideal RCT camouflage ignorance and uncertainty as if it were risk and thereby nurture some of the dangerous control illusions of financial engineering whose consequences we experienced in the financial crises.

philosophy and science as rational endeavors (upper branch in graph 1) rejects the traditional rationalist prima philosophia claim of ruling over science in a manner impervious to scientific knowledge and criticism.[4]

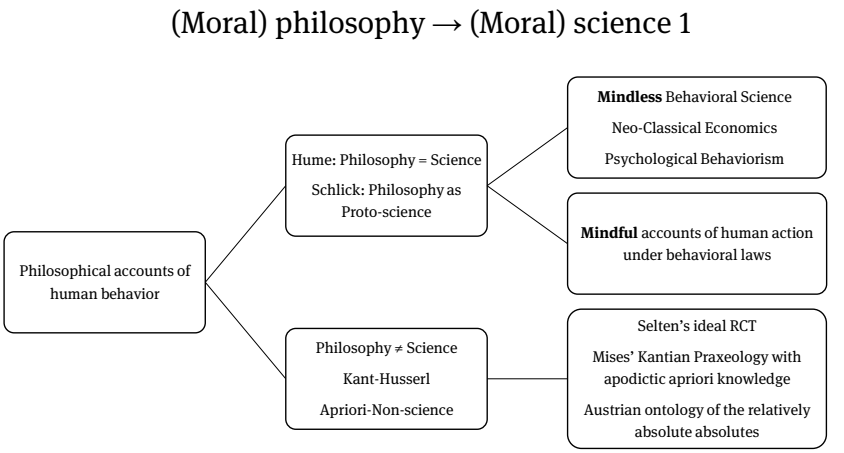

Fig. 1

To the extent that ideal RCT implies such a claim it is subject to basically the same criticisms as classical rationalism. I will not engage the discussion of classical rationalism here but rather focus on Reinhard Selten's ideal theory of interactive rational choice making which avoids classical criticisms by categorically separating RCT from disciplinary science (as a system of broadly 'experiental' insights). The ideal rational choice theory part in Selten's methodological dualism does not make any of the traditional rationalist claims of characterizing a priori what the scope and limits of a posteriori considerations may be. Since it also denies the Humean continuity between philosophy and science it, however, raises questions concerning the role of such an ideal RCT.[5]

4 Beyond the claim of a foundational unity of philosophy and science I follow Moritz Schlick 1986 in regarding it as a—if not the—primary task of philosophy to prepare fields of intellectual interest such that they can be handed over to a disciplinary evidence-oriented treatment (i.e. a science in the present disciplinary sense). Schlick took his inspiration from the psychologist Oswald Külpe (Külpe 1897); see in particular chap. IV, §31, problem 3.

5 Many of the so-called Austrian, in particular von Mises type, economists seem to endorse a methodological dualism with such stronger rationalist claims; see Kliemt 2017.

It is not accidental that Selten refers to ideal RCT as 'rationology'—in analogy to theology. The perfectly rational cognitively unlimited individual envisioned in ideal RCT is akin to the infinitely powerful, benevolent and knowledgeable being(s) studied by ('natural') theology. In the rationological study of ideal RCT infinitary assumptions lead to the same discontinuity with worldly phenomena (see *section 4*) as infinitary conceptions in case of the envisioned deity. In both cases issues of internal coherence as well as the credibility and origins of 'assumptions' emerge as central.

Undoubtedly rationology is a potentially fascinating intellectual field. Yet, its role may be as problematic as that of theology. It invites interpretations of approximability and explanatory power which in fact are non-existent.[6] Like theology, rationology may support practices that must be regarded as harmful in terms of our worldly common sense. And, the empirical evidence concerning many (ab-) uses of RCT seems to suggest that this 'model risk' is quite real.

I shall henceforth focus on the uneasy relationship between science and in some—still to be clarified—sense of those terms 'ideal' and/or 'realistic' variants of RCT. With respect to RCT it is crucial to distinguish between approaches that are at root Humean psycho-logical and those that are Hobbesian decision-logical.[7] The former are experiental in the modern sense of empirical (cognitive) psychology of (boundedly) rational behavior and stand at least in continuity with modern conceptions of evidence-oriented science. The latter are non-empirical in that they proceed in the classical spirit of an a priori analysis 'more geometrico' or some form of analysis akin to it (e.g. Kantian transcendental arguments). Whatever their attractions, they are not in line with modern conceptions of evidence-oriented science.[8]—Before turning to modern ideal theory conceptions that try to clarify a 'normative' ideal of rational choice under the premise that rational choice makers behave in ways compliant with the prescriptions of RCT it is useful to go back to 'square one' and to recapitulate essentials of a decision-logical Hobbesian approach to what traditionally were dubbed 'moral subjects' (prescriptive and descriptive).

6 As a consequence of its 'otherworldliness', ideal RCT cannot be treated as a Weberian ideal type approximable by real types.

7 Perhaps my reading of Hobbes is too extreme, see again Külpe 1897, 10. This stylized account is meant to emphasize those traits of the Hobbesian approach that render it a clear pre-cursor of the disciplinary radicalization that later should become economic imperialism but was already recognized by the British Moralists who responded to Hobbes; see Kliemt, 2009 and below.

8 I hasten to add that Reinhard Selten is one of my intellectual heroes because he is a champion of both fields.

3 Pillars of Hobbesian RCT

Spinoza's streamlined endorsement of originally Hobbesian views is characteristic for the history of ideas trajectory along which RCT developed to its present form and inherited most of its foundational problems along the way[9]:

> "Now it is a universal law of human nature that no one ever neglects anything which he judges to be good, except with the hope of gaining a greater good, or from the fear of a greater evil; nor does anyone endure an evil except for the sake of avoiding a greater evil, or gaining a greater good. That is, everyone will, of two goods, choose that which he thinks the greatest; and of two evils, that which he thinks the least. I say advisedly that which he thinks the greatest or the least, for it does not necessarily follow that he judges right. This law is so deeply implanted in the human mind that it ought to be counted among the eternal truths and axioms.
>
> As a necessary consequence of the principle just enunciated, no one can honestly forego the right which he has over all things, and in general no one will abide by his promises, unless under the fear of a greater evil, or the hope of a greater good [...]. Hence though men make promises with all the appearances of good faith, and agree that they will keep to their engagement, no one can absolutely rely on another man´s promise unless there is something behind it. Everyone has by nature a right to act deceitfully, and to break his compacts, unless he be restrained by the hope of some greater good, or the fear of some greater evil." (Spinoza 1951[1670], 203–204)

In view of the preceding citation it seems somewhat strange that economists tend to refer to Adam Smith as the founding father of their discipline. He certainly was the first institutional economist and he shared with his older friend Hume the focus on what nowadays became experimental and (psychological) behavioral economics. However, Hobbes and, even more so, his direct followers like Spinoza, were much closer to the RCT approach endorsed by modern economists than was Adam Smith (and, for that matter, Hume).[10]

9 Spinoza shook off the last residuals of the older tradition that lingered on in the work of Hobbes to endorse a fully 'economic' rational choice account; see on this also Steinberg 2013.

10 Hume endorsed the homo oeconomicus model only as a contrary to fact 'stress test' for moral-political institutions. He suggests "that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a knave and to have no other end, in all his actions, than private interest" (Hume 1985, VI/I, 42). But Hume says, too, that "it appears somewhat strange, that a maxim should be true in *politics*, which is false in *fact*" (Hume 1985, VI/I, 42–43).

The Hobbesian empire of moral science was built on three pillars (that should re-emerge as defining characteristics of the so-called 'economic imperialism' of the second half of the 20th century):

- *First*, theories of human (inter-)action should be derived 'more geometrico' from first 'self-evident' a priori principles of individual behavior ('*a priorism*'). (1.)
- *Second*, overt human behavior is to be explained as serving individual actors' interests in reaching aims, ends or values *('self-regarding individualism*'). (2.)
- *Third*, behavior can be explained exclusively by relating it to (future) causal consequences of each act taken *separately* within *consistent* individual choice-making ('consistent *case by case opportunism*'). (3.)

The three elements show up in Hobbes' original account. The briefest confirmatory passage (1.) is supported in Chap. 4 Leviathan and passim while for (2.), (3.), the briefest confirmatory passage is the first sentence of the central Chap. X Leviathan (Hobbes 1968[1651]) which commences with: "The Power *of a Man*, (to take it universally,) is his present means, to obtain some future apparent good." The present means part is important since it draws attention to power as potential of an individual to reach her aims, ends, or values. This potential is not power over somebody as in the Weberian concept of power. It is a potential "to obtain some future apparent good". Here the 'some', I take it, indicates that there is a pluralism of goods, 'future' implicitly acknowledges the teleological element in human action, 'apparent' makes it clear that human action aims at what is subjectively perceived by the actor as an end.[11]

On the whole, present neoclassical economics still seems to be built on the three pillars of Hobbesian RCT. Individuals' motives need not necessarily be selfish. As rational choice theorists have argued time and again *consistent* pursuit of

11 There is no externally given objective end as in Aristotle. I readily admit that such an end would emerge in an interpretation of Hobbes moving the 'natural obligation' to secure survival center stage. If this medieval element in Hobbesian terminology and thought is taken seriously a completely different type of moral science emerges. It is either leading back to natural law traditions that do not cohere with the body of Hobbes' work—in particular not with his concept of a natural right as *absence* of *any* obligation—or it may lead to an anticipation of evolutionary competition of the Schumpeterian kind. In the latter case one has to take into account the Hobbesian insight that humans need to engage in pre-emptive strikes out of 'defensio'; they are in foro interno under an obligation to hope that this may not be necessary (they understand in modern parlance the Pareto superiority of non-pre-emption) yet in foro externo they are justified to act otherwise; see on this also Kant 1977[1798] who in his metaphysics of morals ß39–43] expresses similar views.

ends according to well-defined complete preference orders suffices to represent human opportunity seeking action *as if* maximizing a well-defined function.[12] To put it slightly otherwise, to unleash the analytical power of RCT—as a 'tidy' theory, in which preference orders can be represented by the natural order of the values of a real valued function—the aims ends or values must be pursued consistently in a self-regarding manner but otherwise may be of whatever content.[13]

Much of the neo-classical tradition in economics has focused on the consistency aspect of rational choice making. It became often oblivious of the fact that at the Hobbesian origin pursuing aims consistently was merely one and perhaps the less important aspect of rational choice as compared to the faculty to act in an opportunity-seeking case-by-case manner that takes into account only the *future causal* consequences of each act taken separately. As John Hicks succinctly put it in his discussion of 'causality in economics': "people would act *economically*; when an opportunity of an advantage was presented to them they would take it." (Hicks 1979, 43)

Spelled out in terms of opportunity-seeking the assumption that 'people would act economically' can be interpreted in at least two different ways.[14] These lead to different conceptions of idealization in RCT: On the one hand, we are informed that human actors will seize opportunities to an extent that allows to build economic models on the assumption that the underlying generalization is as a matter of fact *approximately* fulfilled. On the other hand, we are informed about a presupposition of an ideal theory discussion that intends to develop prescriptions for ideally rational behavior under conditions of universal compliance with the characterization of ideal behavior and the prescriptions or recipes derived from it.

The role of idealizations of the first kind is akin to the role of Weberian ideal types in social and assumptions like frictionless motion in natural science. The role of idealizations of the second kind is akin to Morgenstern's condition of 'theory absorption'—or rather absorbability—according to which it must be possible that the prescriptions of a theory of ideally rational action in situations of interactive choice making hold good if the theory is commonly known and universally observed.

If idealizations in RCT were exclusively of the first, Weberian kind they could play an obviously legitimate role as approximately valid stylized generalizations. The corresponding RCT would be in continuity with science in that it could be

12 As if maximization and fulfilling the appropriate consistency and continuity conditions characterizing rationality amount to the same.

13 See on this clearly and succinctly Hausman 1992, ch. 2.

14 Not as an exegetic question concerning Hicks' work but as a systematic, general one.

critically assessed by means of evidence-oriented science. Contrary to this, idealizations of the second kind need not make any claim to approximate real behavioral facts. The failure of such ideal RCT to live up to some reality test is not decisive. In fixing the 'prescriptions' of what should ideally be done the prescriptions are assumed to be 'descriptions' of what would be done under contrary to fact assumptions.

The ideal theory leg of the methodological dualism of Reinhard Selten is not Weberian. As a philosophically advanced approach to ideal RCT it holds that the theoretical idealizations of ideal RCT are projections of pre-theoretical thoughts about ideal rationality. It intends to assess the speculative projections in terms of their inner logic and under the 'side-constraint' of absorbability of ideal RCT among perfectly rational reasoners with unlimited reasoning capacities. In short, ideal RCT shows how humans would project their intuitive boundedly rational views concerning full or ideal rationality if they were equipped with the mathematical skills necessary to spell out such a projection precisely and completely.[15]— As illustrated next, rejecting the continuity between philosophy and science, Selten's methodological dualism leads along the lower branch to the lowest of the end-nodes of graph 1.

4 Methodological Dualism and RCT

In terms of widely shared stereotypes the two arguably most prominent self-declared methodological dualists in economics, Ludwig von Mises and Reinhard Selten, make strange bed-fellows. Mises and his followers cultivate resentment against RCM (unfortunately identifying it with an old-fashioned conception of 'quantitative methods'). Selten and fellow game theorists endorse RCM and try to push it to its mathematical limits. Yet, Mises and Selten in his incarnation as a full-rationality theorist (his cognitive psychologist incarnation will be addressed below as a different matter altogether) are close to each other in that both are to be placed on the non-science branch of graph 1.[16]

15 The complete separation between ideal RCT and real theory implied by Selten's methodological dualism also implies that large parts of, for instance, economic general equilibrium theory are to be classified not as behavioral science but rather as non-science.

16 The bounded rationality theorist Selten who tries to develop a 'mindful account of human action under behavioral laws' is firmly located on the upper branch of graph 1 and its middle end node.

As a theorist of full rationality Selten avoids the more extreme claims to generate (apodictic) a priori insights about rationality and intentionality that Mises and his adherents tend to raise. Yet, due to Selten's focus on what individuals *think* 'about rational behavior' his approach to full rationality is very close to Mises' praxeology and its Kantian undertones.[17]

4.1 Selten's Ideal Theory

The problem of absorbability of theories—that is, whether theories about human interaction can remain true if the fact is accounted for that the entities whose actions are described and predicted by theories, understand the theories and respond to this knowledge by opportunity taking changes of their intended actions—is akin to familiar ideas about self-affirming and self-refuting theories (prophecies). Through Oskar Morgenstern it found its way into game theory.[18] Here is what leading game theorist Reinhard Selten says about ideal RCT to which he refers as 'normative' decision and game theory:[19]

> "[…] let me explain my epistemological position on methodological dualism. In my view, there is a fundamental difference between normative and descriptive decision and game theory. Normative decision and game theory has the aim of exploring full rationality and its consequences. Full rationality is an ideal about the adequacy and coherence of decision-making. It is not meant to be descriptive of how human beings actually behave, but rather of what they think about the structure of the behavior of an idealized decision-maker without any cognitive limitations. This idealized decision-maker is a mythical hero, whom we may call 'fully rational man'. Real people have limited powers of logical deduction and computation, but fully rational man has instant access to everything that needs to be logically deduced or computed for adequate and coherent decision-making. […]

17 Praxeology tries to develop RCT in the Hobbesian spirit as substantive knowledge a priori (either of the Kantian kind or that of classical Euclidean geometry). Selten and fellow game theorists along with mainstream mathematical economists endorse modern more formalist conceptions of mathematics and axiomatization. Selten is accepted as an important mathematical economist by the present economic mainstream whereas von Mises and his followers are—in view of their interpretation of and resentment against RCM with some justification—sidelined as 'old fashioned', non-orthodox social theorists. Yet there is much more in common between the two main methodologically dualist camps than superficially meets the eye.

18 For more on Austrian influences in the conception of game theory by von Neumann and Morgenstern, see Leonard 2010.

19 From now on I will focus on the Selten position and only occasionally comment on its relations.

> Empirical arguments are irrelevant for normative decision and game theory. What counts is the appeal to underlying tendencies in the thinking about what fully rational man is like. The situation is similar to theology, which is concerned about what we should think about God. […] In view of the analogy to theology, the study of fully rational man may be called 'rationology'." (Selten 1999, 303–304).

The ideal rationality of 'fully rational man' is, as Selten puts it in the citation, "not meant to be descriptive of how human beings actually behave, but rather of what they *think* about the structure of the behavior of an idealized decision-maker" (emphasis added, H. K.).[20] In delivering his stylized account of 'thoughts' concerning ideal rationality Selten relies on two crucial contrary to fact assumptions one explicit and one implicit. Explicitly he states that "fully rational man has instant access to everything that needs to be logically deduced or computed for adequate and coherent decision-making". Implicitly he accepts the constitutive premise of game theory that the concept of 'full rationality' in interactive decision-making is to be developed under the presumption that real behavior complies with ideal RCT.[21]

In terms of familiar philosophical discussions Selten aims to develop a so-called ideal theory.[22] Within Selten's rather precise game theoretically informed framework the ideal theory should be absorbable under ideal conditions.[23] There are no cognitive limitations and the theory of rational play is common knowledge among fully rational actors. The rational actors ascribe rationality as explicated by the theory itself to all actors and behave accordingly themselves presuming that all others do and know that they do etc. Under full compliance with the theory no actor should have a rational reason to behave other than the ('normative') theory of ideally rational behavior recommends under the presumption that the theory is fully complied with by all other actors.

20 As in the case of Herbert Hart's *The Concept of Law*, to explicate 'law' a whole theory must be developed in case of 'ideal rationality' in interactive choice making. Other than in the case of Hart 1961 who wanted to provide a realistic theory built on stylized facts, Selten intends to present an ideal theory separate from behavioral facts; see on 'explication', Carnap 1956, appendix.

21 Examples of the by now extended discussion of 'ideal theory' are Brennan/Pettit et al. 2005; Anton Leist drew my attention to Ypi 2010 as background for the present context.

22 See for a useful (economist's) account of the philosophical discussion Hamlin/Stemplowska 2012, where rationology would presumably be subsumed under "the theory of ideals".

23 The theory addresses all actors by its prescriptions that recommend what they should do and by its descriptions that predict what they would do under certain circumstances.

4.2 Why Selten's RCT Does Not Characterize a Weberian Ideal Type of Rational Behavior

Selten's rationology drives the aforementioned (Hobbesian) philosophical conception of opportunity taking behavior to its extremes. Dynamic choice-making is about exerting proximate causal influences without emotional and cognitive constraints exclusively in a non-myopic view of all future expected consequences of each choice taken separately. Case-by-case opportunism of the underlying teleological action model is exercised without any emotional or other decision-'inertia'. Rational individuals can and—by assumption—will immediately shift their behavioral gears should they perceive an opportunity for exerting a causal influence on the future (serving their 'given' aims, ends or values).[24]

To see why the preceding 'idealizations' create models of RCT that are behaviorally non-approximable consider the central RCT example of the repeated pd, prisoner's dilemma, that is typically taken to show exactly the opposite: by extending the time horizon, ideal model behavior increasingly approximates real behavior. So, imagine an n-times, n>0, identically repeated 2X2 pd basic game pd(n) of two actors i=A,B. Let C_i, D_i, i=A,B refer to the co-operation, respectively, defection move of A,B in the base game.[25] The unique dominant strategy solution of the base game is (D_A, D_B). In the game of length n=1, on round t=1 the move combination $(D_A, D_B)(1)$ will be in equilibrium. However, with '$P_i(t)$' interpreted as 'preferred by i at time t' we have '$(C_A, C_B)(1)\ P_i\ (D_A, D_B)(1)$, i=A,B, that is the equilibrium result is Pareto dominated by the result of mutually co-operative behavior. Note also, if at stage 't=0'—before the game is played—both A and B plan on making a co-operative move the execution of these strategic plans could not be made credible in pre-play communication among fully rational actors who know of each other's full rationality. Consistent joint pursuit of gains clashes with opportunity-seeking behavior: A strategy can be chosen as a plan at t=0 yet the plan to move

24 In particular the familiar generalizing counterfactuals of 'what if everybody would do the same?' and 'what if I would always do the same?' are treated as rationally irrelevant and motivationally impotent since the rational actor knows that fulfillment of the contrary to fact statement is not among the causal effects of any single action. Genuine rule following behavior as well as retributive inclinations (in particular retributive emotions) are ruled out by future directedness of choice in view of the causal consequences of each single act taken separately. Together with the assumption of unbounded cognitive capabilities such counter-intuitive consequences as payoff dominated solutions based on backward induction in centipede and finitely repeated prisoner's dilemma-games can be derived.

25 A few remarks on the formal structures involved must suffice to indicate the basic line of argument. How it would have to be pursued in a fuller account seems rather obvious but would create a lot of notational clutter.

is not the move at t=1. The plan at t=0 can be executed only in the future when the occasion to move (to exert a causal influence) at t=1 arises. Then, however, since with n=1 there is no t>1 the expectation that there are no future causal consequences will dictate—see Spinoza and Hobbes—that the co-operative strategy will not be executed in view of the dominance of defection.

As the work of some collaborators of Selten vividly illustrates,[26] if the human faculty to act opportunistically (the pillar 3 of Hobbesian RCT) is taken seriously all continuity between ideal RCT and real behavior is lost. Accounts of orderly behavior in terms of evolutionary selection will not work if full opportunism is assumed[27]—evolution can only operate on what is invariant rather than on fully flexible behavior—and even so-called Folk Theorems of eductive game theory (Binmore 1987/88) will not 'survive'. The conventional claim that ideal RCT may be regarded as approximately representing central characteristic features of real rational behavior brakes down.

To see why exactly, return to the simple example of the pd(n) but now consider pd(∞) a game identically repeated indefinitely (at least potentially).[28] Then—except for renumbering—the future looks strategically identical, independently of any preceding history.[29] An exclusively future directed analysis of structurally identical expected futures should come to the same conclusion regardless of the preceding history. That is, if the history has been $(C_A, C_B)(k)$ for k=1, 2, …, n the continuation strategy with respect to the future must be the same as after $(D_A, D_B)(k)$ for k=1, 2, …, n and also for any other starting point of a subgame starting at t=n.

An additional argument is required to show what the identical supergame strategy should be. In view of backward induction arguments[30] that single out

26 See in particular Güth/Leininger/Stephan 1991.

27 Only indirect evolution that operates on the rules of the game can be modeled to work while upholding the teleological case by case model of opportunistic action which precludes anything but choosing strategies dominant with respect to the future; see on indirect evolution Güth/Kliemt 2000; Güth/Kliemt/Peleg 1999 or Berninghaus/Güth/Kliemt 2003.

28 The assumption that the end point is stochastic will not help; for, if backward induction is to be avoided, a continuation probability strictly greater than the positive threshold below which backward induction would kick in must apply indefinitely.

29 Think of the natural numbers, if you cut off the first n, you can renumber and start with n+1 as the new first number in the progression which is structurally identical with the original one.

30 If there is a last round n in which defection is dominant, then on round n-1 it must be commonly known among fully rational actors that behavior on round n-1 cannot induce rational actors on round n to behave other than dictated by dominance. Therefore, the dominance of defection on round n-1 will be effective. This is known among the actors on round n-2 … etc. until the all-D behavior emerges.

strategies implying move combinations $(D_A, D_B)(k)$ for k=1, 2, …, n and all finite n, only (D∞, D∞) will be approximable in pd(∞).[31] Under conditions of ideal rationality, contrary to folk wisdom, extending the time horizon to infinity, though avoiding backward induction, will not yield any co-operation as observed in real human behavior.

The upshot of all this is that in a world of fully rational strategic actors who decide exclusively opportunistically in view of substantive pay offs so-called conditional co-operation can*not* emerge. Consequently, the so-called 'order problem' has no plausible solution among fully rational actors who are extrinsically motivated by substantive payoffs.[32] Taking the ideal model seriously co-operative orderly behavior cannot be approximated at all. Since we do as a matter of fact observe human co-operative behavior, strictly Hobbesian ideal RCT becomes an empirical absurdity.

If we assume that ideal RCT correctly identifies and drives to its extreme what we *think* is an essential human faculty—to seize opportunities—we must conclude that ideal opportunistic rationality is otherworldly. Other than in case of Weberian ideal types there is no continuity between the theory of ideally and the theory of realistically rational choice. In Selten's methodological dualism there are no 'bridges' between the ideal and the real theory strands.[33]

If the preceding diagnosis holds good, then whatever the merits of rationology it does not contribute to our experience-based scientific knowledge of the world.[34] To put it slightly otherwise, what we think about the world in terms of ideal RCT does not tell us anything about the world as object of that thinking.[35] Selten is right, ideal RCT is categorically separated from real RCT. In line with this in graph

31 Only so-called subgame consistency according to which structurally identical games should have structurally identical strategic equilibria as solutions is plausible if only the future matters. Strict future orientation rules out all path-dependence of play in structurally identical sub-games and thereby co-operation conditional on past events.

32 Parsons 1968 is a locus classicus concerning the so-called Hobbesian problem of social order.

33 As far as science is concerned Reinhard Selten's dualism is reduced to the monistic psychological founding of rational choice modeling in terms of so-called bounded rationality which Selten himself has always propagated. As an activity outside of science the a priori reasoning of pure rational choice theory may play a role.

34 The other forms of a priori reasoning represented in the same end node box will not be discussed here since the focus is on RCT in its present mathematically sophisticated form not burdened with the epistemic and ontological liabilities of the traditional variants.

35 Making their distinction between how the world is perceived and how it is 'as such' even Kantians would concede this, yet insist that something can be known a priori about the world as an object of experience.

1 the lower branch on which Selten's ideal RCT is located is separate from science. The continuity between science and philosophy is interrupted.

Selten—like many Austrian economists and broadly Kantian scholars of social theory—allows for forms of rationality that transcend the rationality of science. Where in particular Austrian economists mix their claim that a meaningful fact-insensitive conception of intentionality and rationality exists with a (mis-) conception of the limits of mathematization, Selten's discontinuity claim arises from a consequent application of mathematical reasoning to teleological forward looking choice making. Both camps believe that there is a realm of meaningful ideal RCT beyond human practices that are conventionally regarded as rational (as in particular those of empirical science). In Selten's case there is even an argument akin to a proof that ideal rationality as emergent from a projection of human intuitive conceptions of rationality and real rationality fall apart.

The thesis that the human ability to make choices teleologically puts human behavior beyond (causal) explanations based on law-like regularities—at least at the present state of empirical knowledge—is independent of any limits of logic and mathematics per se. As in particular the example of Selten's methodological dualism shows it deserves to be taken seriously. It cannot be brushed aside by other adherents of mathematical social science as a manifestation of the resentment of those who feel left behind due to their lack of mathematical skills. In particular the critics of 'internalization of values' and other concepts akin to intrinsic motivation and purely subjective factors in choice making will get into trouble if they take their own rhetoric seriously. Their a priori preference for behavioristic explanations in terms of extrinsic motivation is within an evidence oriented conception of scientific rationality unconvincing as long as they cannot beef up their claims by presenting superior explanations in terms of 'mindless RCT'. The alternative to this kind of approach is a cognitive psychology version of 'mindful RCT'. Both are incoherent with ideal RCT.

To put it slightly otherwise, what is needed is an account of human rational choice making that either avoids teleological concepts altogether (as in behaviorism) or embeds them in an empirical conception of teleological reasoning (as in cognitive psychology). Some concept of 'rational choice' may be explicated in terms of overt behavior and its 'successes' in (inter-active) choice making. A 'mindless RCT' that explains behaviorally 'rational choice' in terms of causal factors other than 'reasoning' (including in particular selection and adaptation in competitive processes) should be explored. There is room for this, on the upper branch of graph 1. However, there is also clearly room on the upper branch of graph 1 for 'mindful RCT' which links rationality psycho-logically rather than logically to reasons and reasoning.

5 The Nature and Significance of RCT

Mindful RCT acknowledges as a brute fact of human self-experience that real human actors command a basic faculty to distinguish between acting/intervening (teleologically) into the course of the world and representing/predicting the course of affairs under (causal) laws. An intentionally acting human being will perceive herself as ('self-controlled') *author* of her actions. She will imagine herself as being able to reverse her intentions at any point in time. She is to herself a maker rather than a predictor of her choices.[36] Whether or not she regards herself as rationally or morally justified she will think of herself as intentionally bringing about future causal consequences.

From an internal point of view, in a 'first-person' perspective, making a choice is incompatible with predicting it. Yet this does not rule out predicting choices from a point of view external to the choice maker. Humans can do better than chance in predicting choices of fellow humans.[37] It seems that there can be predictive and explanatory RCT while accepting the first-person agent-relative *thinking* about choice making as separate from choice prediction.

The explanatory subjectivism of mindful RCT accepts the presence of other mindful choice makers as a fact. It conjectures that to emulate or to reconstruct the internal point of view to the action situation may be helpful for external explanatory and predictive purposes. Whether or not this in fact is true depends on whether or not reliable law-like hypotheses exist. Here as in other fields prophecies turn into evidence based predictions to the extent that they can be based on tested and corroborated law-like (stochastic) hypotheses.[38]

Within a realist conception of RCT, law-like hypotheses can themselves not be merely a matter of thinking about the world. They are so to say not only 'in the expecting mind' of an observer but are conceived by the observer to operate on observed actors as other 'laws of nature'. In a 'mindful' psychological science

36 This will hold good even if the actor knows that according to results of neuroscience her decisions have been fixed in her brain before they become conscious to her. On the surface of conscience she will experience herself as a 'self-propelled' actor 'pulled' by the expected future rather than being pushed by the past. Even if she is driven by retributive emotions of a positive or negative kind and is conscious of it there will be the 'illusion' of commanding the faculty to act otherwise.

37 To what extent predictions from an external point of view to actors are possible is an altogether contingent empirical question; see on empirical evidence concerning the scope and limits of prediction in human affairs, Tetlock 2009; 2015.

38 Like those identified in the heuristics and biases approach; see Kahneman 2012 for a popular discussion.

approach to human behavior the observer will typically pre-suppose that law-like regularities relate subjective perceptions of action situations (stochastically) to overt behavior of actors.[39]

If an observer uses his reconstruction of a 'subjective' action situation of another actor to predict the behavior of the other individual according to objective law-like hypotheses concerning cognitive processes then—despite the reconstruction of the subjective situation—he adopts what has been called an 'objective attitude' (Strawson 1962). Such an observer might use objective law-like regularities to rationally 'manipulate' the actions of another rational individual in a purely causal way. If the observer feeds back his information to the actor then the actor can respond in a possibly 'falsifying' way in rational pursuit of her own ends (as, e.g., in case of self-refuting prophecies) yet this alone does not change the situation into one in which the observer approaches the actor with what Strawson called a 'participant's attitude'.

It is obviously hard to draw the relevant distinction between an argument offered from a participant's and one used for manipulation from an objective point of view precisely.[40] An argument may be used for the sake of its causal effects or for the sake of its internal validity and often for both reasons simultaneously.[41] In any event, 'subjectivism' as methodological precept of taking into account subjective mental states as well as mental models of action situations in explanatory RCT efforts is compatible with science in general and a 'technological' conception of the application of scientific knowledge.

What individuals think about what is rational according to RCT can matter as a causal factor. For this it needs to enter mental models of real subjects as a real factor. In principle the corresponding RCT remains a branch of (cognitive) psychology.[42] That RCT may be known to the individuals who are described by it

39 The logic of a situation can be very compelling but empirically a law-like regularity that individuals will act according to what they perceive as implied by that logic is needed for a fully-fledged explanation.

40 In a related context this distinction is more carefully discussed in Baurmann 1987. Whether an act of manipulation or of, say, genuine agreement seeking occurred is contingent on the aims, ends or values of the external actor. He can address 'arguments' towards 'another mind' with the *intention* that the other should 'own' and thereby ratify them or as causal forces meant to bring about predicted responses.

41 Without a 'normative' fact that somebody as a matter of fact intends to pursue certain aims, ends or values there is no instrumental or technological 'ought'. Which aims, ends or values as a matter of fact prevail (are 'given' in the sense of Robbins 1935) decides to which technological uses knowledge will be put.

42 The law-like regularities underlying cognitive psychology accounts are beyond human intervention as much as those invoked by ('old fashioned') psychological behaviorism.

does not in itself change its explanatory nature. However, it will make predictions more and sometimes exceedingly complicated.[43]

Psychological explanatory behaviorism that strives to explain overt behavior exclusively in terms of regularities in overt behavior (uppermost end-node of graph 1) would be a nice way out of the problems of unpredictability arising from symmetric knowledge of RCT.[44] Law-like regularities that could explain regularities in overt behavior without taking recourse to the directly non-observable theoretical concepts of cognitive psychology would be methodologically preferable for other reasons as well. However, the condition that the relevant explananda of (inter-active) choice making can—at least approximately—be explained in behavioral terms will typically not be fulfilled. It seems that the regularities in overt behavior that we as a matter of fact observe and desire to explain require taking recourse to cognitive psychology concepts like in particular genuine rule following behavior by a choice making actor from the actor's internal point of view.

To the extent it is required for and conducive to explanations explanatory subjectivism seems unavoidable. It is based on the presumption that the laws of nature that govern rational choice relate to subjective perceptions and cognitive processes. Yet the law-like regularities are assumed to exist 'for real' and not only as a logic of situations etc. (locating such RCT at the lower end node of the upper branch of graph 1).

The explanandum of the factual existence of social order that was used in the preceding section to illustrate that ideal RCT will not work as adequate explanans may serve as an illustration here. The game analyzed in RCT terms arises from a game *form* that exhibits pd like material or substantive payoffs as extrinsic motives or incentives. It is a pd in terms of directly observable substantive 'pay offs' but not necessarily in subjective payoffs. The subjective payoffs may be quite different from the objective or substantive ones. Often, in a 2X2 pd game form individuals i=A,B will form 'subjective' preferences according to which not only $(C_A, C_B)\ P_i\ (D_A, D_B)$, i=A,B but also, $(C_A, C_B)\ P_A\ (D_A, C_B)$ and $(C_A, C_B)\ P_B\ (C_A, D_B)$. Then the unique dominant strategy solution of the base game is not anymore (D_A, D_B). If

43 See for an exploration of so-called theory absorption among boundedly rational actors, Güth/Kliemt 2004 and on theory absorption more generally and precisely Dacey 1976; 1981, building on Morgenstern.

44 As for instance the example of finance shows rational choice makers who know RCT and other theories about the world and know that they know will engage in theorizing in symmetric ways that will render the predictive efforts self-refuting and the so-called random walk down Wall Street may be the result. Reasoning about knowledge seems as a matter of fact influencing the world.

(D_A, D_B) P_A (C_A, D_B) and (D_A, D_B) P_B (D_A, C_B) a so-called AG, assurance game, arises from a pd game form in substantive payoffs due to subjective factors.

This is merely one of the better known examples for the necessity of including subjective factors in adequate explanations of (inter-active) rational choice making.[45] It obviously generalizes to the claim that RCT can have adequate explanatory force only to the extent that it is a variant of cognitive psychology (focusing on intentional human behavior in inter-active choice making). For, the subjective perceptions of situations leading to games and not only the game forms matter. In particular economists are prone to point this difference between game and game form out. However, they then take resort to their conventional strategy of 'let us assume that the game form and the game motivationally coincide' which avoids all the interesting and crucial empirical issues of the relationship between games and game forms in realist RCT.

Under what conditions individuals who are exposed to a pd game form are in subjective preference terms interacting in games of the PD or AG type is an empirical issue. Our ability to answer the related empirical questions on the basis of nomological knowledge will be decisive for our ability to predict behavior. Yet, however this may be, descriptive explanatory RCT in both, its cognitive psychology and its behaviorist incarnation, remains firmly located on the upper branch of graph 1 (regardless of the mindless or mindful nature of explanations).

Though I intend to stand by this account of the descriptive-explanatory nature and significance of RCT I readily admit that RCT has always been closely associated with aspirations to make 'better' choices. These aspirations reach from improving the prospects of getting one's way in pursuit of self-regarding aims, ends or values to realizing broadly speaking moral standards of showing interpersonal respect in interactive choice making (as an 'ethical' extension of the participant's attitude). An account of the nature and significance of RCT would be thoroughly incomplete without at least a few remarks relating such 'normative' RCT concerns to descriptive-explanatory RCT. In doing so one has to bear in mind in particular that Selten referred to rationology also as a 'normative' theory (though not one that could be used to derive prescriptions for real behavior).

45 As opposed to games that represent the preferences over plays of the game (all things considered), game forms represent only the material or substantive paths and outcomes. In particular, a game form of a pd may in fact typically give rise to an assurance game with no dominant strategy in preferences regardless of the fact that it exhibits a pd structure in material or substantive payoffs; a fact to which Amadae 2016 rightly draws attention though due to her lack of the category of a game form her presentation and understanding seems somewhat deficient.

6 RCT as a 'Normative' Account of Human Behavior

Awareness that ideal RCT is not an explanatory scientific theory makes it harder to claim the prestige of science for normative purposes of prescribing what rational actors ideally should do. In a certain sense the 'rational' in real RCT has, however, as a matter of fact normative undertones in the wide sense in which, say, maxims of prudence, recipes expressing 'technological know-how' and standards of 'correct' measurement also are normative.[46] In our search for a WRE on the theoretical and practical merits of RCT we need to account for the normative uses of RCT as well. An answer to this challenge would require a whole book. This being acknowledged the subsequent remarks on WRE present merely some essential demands that more complete and possibly more convincing answers to the challenge of explicating the normative aspect of real RCT would have to meet.

6.1 WRE on Descriptive and Prescriptive Claims

The term reflective equilibrium stands for systematically seeking a coherent account of general and specific claims which may in principle be of a prescriptive or descriptive, an a priori or an a posteriori nature. The search is 'wide' if background theories (which themselves may have been subject to separate equilibrium considerations) are included in the process.

Without going into details and without making any claim to exegetically represent the views of Rawls or Daniels with whom the WRE concept is normally associated in philosophy the basic intuition may be presented the following way:[47] When humans try to systematize and to critically assess bodies of knowledge or advice, coherence is the guiding aim. Typically, specific claims representing specific cases in which observations of what is true or false and also prescriptive or evaluative intuitive judgments of what is right, wrong, better or worse are '(relatively) prioritized' in that the burden of proof is assigned to those who intend to overrule them as invalid when they do not cohere with general or abstract claims.

To give a most simple example, assume somebody looks out of the window observing the killing of another human being. He considers this as repugnant. Assume also that he never thought about such killings before. Even such a single

46 See on what may be called minimum normativity of standard-setting the special issue of *Analyse & Kritik* 2016 (38), and the account of and examples of norm justification in WRE given there.
47 See Rawls 1951; 1971; Daniels 1979. Goodman 1983 whose views are closer to RCT considerations in substance and form will be discussed towards the end this section; see for a succinct account Hahn 2004.

strong situational encounter may induce him to conclude that the general claim that the killing of other human beings is allowed must be rejected. Let us assume also that after further considerations of a general as well as additional observations of a particular kind he feels assured that killing of fellow humans is in general wrong (and stops there in a temporary equilibrium of his reflections since for the time being he sees no counter argument).

After a while the person is confronted with an attack of a robber armed with a baseball club. To prevent becoming a victim, the attacked considers using his gun. To do so would, of course, violate the general principle that 'thou should not kill!' that he came to endorse. Coherence will be reached once the principle is modified to demand 'killing is forbidden unless in self-defense'. New reflections may be necessary if a robber attacks with a toy pistol or, say, with a banana. Depending on often subtle circumstances, self-defense might not justify the use of the gun in reflective equilibrium in such cases.—So much for the toy example which should be sufficient to illustrate the guiding idea.

If a reflective equilibrium is sought on 'great questions' like what are the principles systematizing our views on justice or how to account for our practices of induction (corroborating or refuting claims) systematically, a simple stylized process as described before will not work. Without going into details a few remarks may be sufficient to sketch some of the necessary basic additions.

First, the simple linear process characterized before will not apply. General principles may become so well entrenched that particular claims that contradict them may be discounted in favor of general and abstract principles. This is tricky business since the hurdles must be high before some 'refuting' evidence may be discounted say in case of a general law-like hypothesis of natural science. Not by chance do we have additional rules like the demand that it must be possible to replicate the falsifying phenomenon in a reliable and predictable way etc. These additional rules will normally be sufficient to neutralize fake evidence. Still, clashes between specific claims and general evidence are not rare but very common. For instance, in cases like evidence-based medicine the intuitions of doctors that stem from their personal experience often clash with general evidence-based rules or recipes of good practice. Though there are rather well supported secondary principles that would guide rational doctors to discard their own specific intuitions in cases where, say, the evidence of randomized controlled trials speaks against their specific contrary experiences some of them would still rather follow the testimony of their particular case-based experience. They could often point to arguments that indicate that their cases are special and therefore the general conclusions of the randomized trial do not apply. Nevertheless, the results of

randomized trials should carry much weight in light of the empirical evidence on how randomized trials fare as compared to expert judgment in general.[48]

A second and perhaps even more important point about addressing the 'great questions' in a search for WRE is their relation to established practices. The 'great questions' do not arise out of the blue. Specific claims arise in a context of real (smaller) problems. Against this background the implicit empty slate premises raised by claims to fact-insensitive intuitions of pure reason—or what have you—seem rather absurd. We must start from a contingent state of mind and knowledge when we search for a WRE and what arises will be dependent on the starting point. The vision of path-independent answers to the 'great questions' seems—at least to me—absurd.

More concretely, take Rawls' intention to reach a WRE on fundamental principles of justice. Whatever Rawls and many of his followers may have also said in favor of path-independent claims, his original inclination—which he never fully abandoned—to conceptualize the WRE search as relative to the experiences made in a Western society under rule of law and operating under conditions of intermediate scarcity seems the only plausible one—even if in a world of plural values and diverse cultures the Rawlsian search for an overlapping consensus may not succeed at all. In any event, to systematize particular realms of political convictions will remain a worthwhile endeavor.

Whether or not one agrees in substance with Rawls, his focus on questions of priority—as in his 'priority of liberty'—in the WRE search for a Western 'rule of law society' seems intuitively right (at least to Westerners like me). Yet, when acknowledging path dependence, granting priority to certain claims vis a vis others will be convincing only relative to some contingent base and some value judgments.

The acknowledgment of the relativity of the results of WRE search to a context and a starting point is as essential for other 'great questions' than those addressed by developing 'a theory of justice'. The 'great question' of developing a theory of rational choice is at least as fundamental if not even more so, yet as dependent on contingent facts. As a paradigm of a reflective equilibrium search in such cases Goodman's discussion of induction may be used. Goodman very strongly emphasizes the role of entrenched practices and in that sense 'relativity' to factual context, relevance of a starting point and path dependence. Intuitions formed on an empty slate are sidelined as of minor relevance.

That Rawls thought of claims that concern merely fictitious situations originally as irrelevant should be noted, too. In case of RCT, claims derived from such

48 Ever since Meehl 1954 such second order evidence has been considered; see in a different context impressively Tetlock 2015.

fictions are at least of lesser force than those derived from real cases. The fiction of a decision maker who has unlimited reasoning capacities may trigger certain intuitions in us yet this 'intuition pump' will create a flow of specific claims that is of minor relevance as compared to responses to real decision situations. It cannot completely be ruled out—and certainly not on a priori grounds—that there is some legitimate role for ideal a priori theory in indirectly shaping our WRE—as there is a role for thought experiments as opposed to real experiments—yet it seems that the relatively higher weight or the priority of claims should be with those derived in real situations (the same way as real experiments trump thought experiments, say, in physics).[49]

6.2 Weighing the Evidence in Search for WRE

In search for an equilibrium, trade-offs between what has to be given up and what can be maintained as part of a coherent system of mutually supportive claims must be made. All claims are treated as revisable. Yet, there are—at least for the 'time being'—certain fundamental principles guiding the process of weighing claims. For instance, we would on the most fundamental level not accept as 'minimally rational' the elimination of a well-entrenched (coherent with and supported by many other claims) factual claim simply because it expresses an inconvenient truth. In a similar vein, we would accept that 'ought pre-supposes can' as bridging the gulf between the descriptive and the prescriptive in ways that allow the elimination of claims that demand the impossible. Finally, and perhaps most importantly, in an adequate equilibrium search claims that represent time honored human practices have special weight. Yet again this does not mean that they cannot in principle be revised and eliminated in a reflective equilibrium. It cannot be ruled out a priori that intuitive particular or general claims that are not supported by practices can successfully discharge the burden of proof against claims supported by practices and established institutional facts.

In any event, the preceding tentative rules of weighing the evidence and making trade-offs when seeking for a wide reflective equilibrium give a certain precedence to the factual as opposed to the counter factual.[50] In particular, the acceptance of factual claims cannot directly be overruled by evaluative ones. If we have

49 So-called trolleology is another case in point similar to rationology in many respects. It may be noted, though, that trolleology can give rise to interesting considerations concerning causal relations to brain activity; see on this Greene 2013.

50 Merely imagined possibilities, prescriptive demands that something that does not exist is to be brought about…

good reasons to believe that x is, but good reasons to desire that non-x should be the case the latter should not affect our factual believe in x. If it comes to the relation between representing and intervening the priority is with representing and explaining. Interventions are planned and executed on the basis of factual knowledge—or at least this should be so.

The WRE-search sets in where disciplinary science and explanatory RCT end. Without turning to 'fact-insensitive' ideal RCT the search for WRE is emulating empirical science efforts to reach coherence of general and singular 'statements' for purposes of assessing 'normative' matters. In doing so the search incorporates 'witness-evidence' or 'intuitions' in ways similar to how the testimony of scientists is used in the critical assessment of scientific theories. Yet concerning normative claims the testimony of witnesses is non-observational in that it is not merely used to describe what witnesses *think* about what is 'normatively adequate'. The testimony of 'representative individuals' is rather used directly as a non-descriptive normative premise to support normative standards or rules of correct conduct. As Nelson Goodman states in *Fact, Fiction, and Forecast* (1983, 63 f., cited after Hahn/Schlaudt 2016, 314):

> "How do we justify a deduction? Plainly by showing that it conforms to the general rules of deductive inference. [...] Yet of course, the rules themselves must eventually be justified. But how is the validity of rules to be determined? Here again we encounter philosophers who insist that these rules follow from some self-evident axiom, and others who try to show that the rules are grounded in the very nature of the human mind. I think the answer lies much nearer the surface. Principles of deductive inference are justified by their conformity with accepted deductive practice. [...] A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement lies the only justification needed for either."

If the role of established practices along with de facto singular convictions is regarded as decisive as in Goodman (1983) and in Rawls' (1951) *Outline of a Decision Procedure for Normative Ethics* the dependence ('relativity') of normative justifications on factual convictions, aims, ends and values is acknowledged. Reflective equilibrium contains a stylized account of practices that as a matter of fact are prevailing. Yet the aspiration of the reflective equilibrium search goes beyond evidence based empirics of prevailing practices and the *opinions* of what is and what is not compliant with the 'spirit' of such practices. There is a clear view that the reconstruction of what people would basically deem appropriate—according to the de facto standards of practices and the 'intuitions' shaped by them—can after critical assessment and 'extrapolation' yield general standards of what in a 'local' sense should be done in particular instances.

With respect to developing prescriptive standards of rational action against the background of descriptive RCT it should be noted that the process is sensitive to the facts of general practices and specific judgments that accompany them. The critically rational account of these 'normative facts' yields a concept of what is to be deemed rational as opposed to what is rejected as irrational. It is dynamic in the sense that the self-applicable 'technological' concept of rationality may develop through time. It is like a living constitution of a community of mindful actors who mutually recognize each other as 'members of the club of boundedly *rational* actors'.

The underlying theories need not classify all intentional behavior as rational—qua intentionality as in Mises' methodological dualism—nor is it necessary that they rely on an ideal theory of rational choice—as in Selten's methodological dualism. They specify what is rational in a self-selection process. Such bootstrapping may seem unsatisfactory in many regards yet in view of the list of preceding rules of thumb it is at least not completely arbitrary.

6.3 WRE on Normative Standards of Morality

The broad discussion of the role and legitimacy of ideal theories in philosophy that was triggered by a certain interpretation of the WRE search in Rawls' (1971) *A Theory of Justice* will not be discussed in any detail here. Suffices it to note that the role of a fact insensitive ideal theory that Rawls himself seems to allow for in his exploration of 'justice' is not in line with the contextual interpretation of the WRE search in Goodman's and in Rawls' own earlier account.

That philosophers who reject the continuity of philosophy and science and in particular that of ideal RCT and boundedly rational practice were eager to jump on the ideal theory train of Rawls' theory of justice is unsurprising.[51] Without claiming to offer the only correct interpretation it seems to me, however, that an interpretation of Rawls' WRE-search in 'a theory of justice' is possible in ways that are in line with the continuity view of science and philosophy and the principles of contingent path dependent equilibrium search enunciated before. Then the Rawlsian search for WRE in developing his theory of justice is framed as anchored in

51 The ideal theory interpretation of Rawls is supported by the fact that Rawls uses in his WRE rational choice models for justifying moral principles formulated in ideal RCT terms. Yet according to the argument concerning the nature and significance of RCT offered here this will not work in favor of ideal theory since ideal RCT itself is rejected in the account of rational practice offered here.

stylized accounts of socio-political practices of well working liberal Western legal orders under conditions of 'intermediate scarcity'.

The principles that emerge after several circles of the search and equilibrating process can at best make a claim on those who endorse the underlying practices in general. Such normative standards may be regarded as "relatively absolute absolutes" (Buchanan 1999, vol. 1, 442–454). They are 'absolute' only relative to some contingent facts. In that they are very similar to what Goodman required in giving an account to inductive practices. For the present purposes it seems sufficient to note that a WRE on moral standards will be sought for along the same lines as that on any other standards.

7 Concluding Remarks

Selten's ideal RCT systematizes what 'we' *think* about what is rational in search for a reflective equilibrium. The basic intuitions that serve as springboard for this search are not found by an empirical study of what 'we' are as a matter of fact thinking, though. In his search, Selten conducts no high-level opinion poll, no fact-finding mission concerning normative facts like established rules. Neither are there elaborate descriptions of practices that might serve as paradigm examples of what is conventionally regarded as rational according to human practices of judgment. Quite to the contrary the initial 'intuition material' is 'laundered' and driven to its extremes in an ideal theory equilibrium search in which empirical issues of feasibility of behavioral recipes *do not* figure at all. As in the ideal theory interpretation of Rawls' theory of justice these aspects are eliminated by the a priori assumption of behavioral compliance.

In a Selten type WRE the real practice dimensions would be eliminated to a degree that alienates it from science. What Selten calls normative RCT appeals to reason and aesthetic feelings as do certain mathematical theories. It spells out what the human imagination may fabricate as ideal rationality. According to the Humean view of this paper, it seems that such ideal rationality theory is too far removed from actual human practice to have real relevance. It does not respect principles like ought pre-supposes can and it does not even try to uphold continuity of its substantive prescriptions as rooted in pre-equilibrium intuitions shaped by human practices. Neither is it a Weberian ideal type representing reality in a stylized yet approximable way.

As the experience with the ascent of evidence-based medicine shows, discriminating between practical prescriptions that are merely speculative and those for which tested evidence exists may be expected to lead to an improvement of

our practices. Though it is completely open to what extent all practices may be evidence-based, arguments that assume away the relevance of the evidence-base seem unconvincing. Ideal RCT that simply assumes away empirical limits to RCT will not do as a substitute for evidence-based theories of what we deem rational behavior. The demand for an a priori conception of rationality that pre-cedes all practices (including what we as a matter of fact deem desirable in actor relative ways) is no more such a theory than hunger is bread (see Bentham 1843) or that the demand for an ideal moral theory will provide that theory.

Strict RCT leads to models that cannot explain since the underlying behavioral assumptions are not approximately true. Despite this discontinuity with real behavior, models based on RCT invite ascribing explanatory power to them even in cases where they are not even a far cry of the real causal mechanisms of interactive human choice behavior. By their inherent misrepresentation of the behavioral facts, models based on ideal RCT tend to suggest mistaken technological recipes leading to misguided interventions into real interactive processes. This model risk can be avoided in principle by Selten type strict methodological dualism which insists on the separation of pure a priori RCT and behavioral science a posteriori. Yet the price for this is that the relationship to science becomes discontinuous and the rationality embodied in science and its practices cannot be brought to bear on the transcendental (v. Mises) or mathematical (Selten) a priori speculations of RCT.

As a way out, this essay indicates that going back to Nelson Goodman's and John Rawls' original ideas on reflective equilibrium search in a non-a priori manner may be useful. Goodman dealt with the problem of inductive reasoning by relying on real human practices and trying to systematize them in a critically rational way open for all sorts of argument but insisting on some primacy of real human practices. Rawls' original 'outline of a decision procedure for normative ethics' along with his—not always consistent—insistence that the reflective equilibrium we can meaningfully search for will always be rooted in particular practical experiences under particular circumstances located in time and space is another case in point. How exactly in such contingent wide reflective equilibrium searches scientific results and scientific and non-scientific value judgments should be brought to bear on each other is itself subject to a reflective equilibrium search on the reflective equilibrium search 'method'. Not much about this can be said except that it would seem highly desirable and that 'absorbability' of some kind would seem a desirable property of such a theory of WRE.

That is, almost tautologically, such a conception should not subvert its own prevalence as an equilibrium conception if it spreads. If the continuity thesis is taken seriously then ('ultimately') the absorption of such theories could itself be scrutinized in broadly empirical terms.

Acknowledgment: Alluding to Robbins 1935 by the title of what started as a contribution to the 'Ethics and Economics' conference at the Singapore University of Technology and Design (SUTD) January, 19–20, 2017 is, of course, intended; I am grateful to conference participants for their comments, in particular, Dan Hausman and Zsombor Meder. Extremely helpful were later critical discussions by Michael Baurmann and Anton Leist who induced me to focus on RCT as an ideal theory and its relation to science rather than on 'moral science' more generally. They rightly insisted that I be more explicit on the distinction between Max Weber's ideal types and Selten type ideal RCT and my 'old fashioned' use of the reflective equilibrium concept as expressive of the 'methodological' continuity of philosophy and science. If I am wrong, I will at least be more clearly wrong after their intervention. Sarah-Lea Effert provided some very helpful specific discussion of ideal theories as did the participants of a conference in honor of Alan Hamlin at the university of Manchester a few years ago. All this made me aware of my still insufficient grasp of an extended discussion and literature concerning ideal theory to which I here add, if anything, a critical rationalist 'Albertian', Albert 1985, account of the perspective of ideal RCT which in a way may be the most fundamental ideal theory of all. Last but by far most importantly I should like to mention that my thinking on approximate explanations has been fundamentally influenced by discussions concerning idealization and approximate explanation with a younger Albert. See Albert, M./H. Kliemt (2017), *Infinite Idealizations and Approximate Explanations in Economics*, MAGKS Working Paper, URL: https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/paper_2017/26_2017_albert.pdf.

References

Albert, H. (1985), *Treatise on Critical Reason*, Princeton

Amadae, S. M. (2016), *Prisoners of Reason: Game Theory and Neoliberal Political Economy*, reprint edition, New York

Baurmann, M. (1987), *Zweckrationalität und Strafrecht*, Opladen

Bentham, J. (1843), *Anarchical Fallacies*, vol. 2 of Bowring (ed.), works

Berninghaus, S./W. Güth/H. Kliemt (2003), From Teleology to Evolution. Bridging the Gap between Rationality and Adaptation in Social Explanation, in: *Journal of Evolutionary Economics* 13, 385–410

Binmore, K. (1987/88), Modeling Rational Players I & II, in: *Economics and Philosophy* 3 & 4, 179–214 & 9–55

Brennan, G./P. Pettit (2005), The Feasibility Issue, in: *The Oxford Handbook of Contemporary Philosophy*, ed. by Frank Jackson/Michael Smith, Oxford, 258–279

Carnap, R. (1956), *Meaning and Necessity*, Chicago

Gul, F./W. Pesendorfer (2008), The Case for Mindless Economics, in: Caplin, A./A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook*, Oxford–New York, ch. 1
Daniels, N. (1979), Wide Reflective Equilibrium and Theory Acceptance in Ethics, in: *The Journal of Philosophy* 76, 265–282
Geuss, R. (2008), *Philosophy and Real Politics*, Princeton
Goodman, N. (1983), *Fact, Fiction, and Forecast*, 4th rev. ed., Cambridge/MA
Greene, J. (2013), *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, New York
Güth, W./H. Kliemt/B. Peleg (1999), Co-Evolution of Preferences and Information in Simple Game of Trust, in: *German Economic Review* 1, 83–110
— (2000), Evolutionarily Stable Co-Operative Commitments, in: *Theory and Decision* 49, 197–221
— (2004), Bounded Rationality and Theory Absorption, in: *Homo Oeconomicus* 22, 521–540
— (2007), The Rationality of Rational Fools, in: Peter, F./H. B. Schmid (eds.), *Rationality and Commitment*, Oxford, 124–149
—/W. Leininger/G. Stephan (1991), On Supergames and Folk Theorems: A Conceptual Analysis, in: Selten, R. (ed.), *Game Equilibrium Models. Morals, Methods, and Markets*, Berlin, vol. 2, 56–70
Hahn, S. (2004), Reflective Equilibrium: Method or Metaphor of Justification?, in: Löffler, W./P. Weingartner (eds.), *Wissen und Glauben*, Wien, 237–243
—/O. Schlaudt (2016) (eds.), Logic, Morals, Measurement—Origins and Justification of Norms, in: *Analyse & Kritik* 38, 311–316
Hamlin, A./Z. Stemplowska (2012), Ideal Theory and the Theory of Ideals, in: *Political Studies Review* 10, 48–62
Hart, H. L. A. (1961), *The Concept of Law*, Oxford
Hausman, D. M. (1992), *The Inexact and Separate Science of Economics*, Cambridge
Hicks, J. (1979), *Causality in Economics*, Oxford
Hobbes, T. (1968[1651]), *Leviathan*, Harmondsworth
— (1990[1682]), *Behemoth*, Chicago
Hume, D. (1985), *Essays. Moral, Political and Literary*, Indianapolis
Kahneman, D. (2012), *Thinking, Fast and Slow*, London
Kant, I. (1977[1798]), *Die Metaphysik der Sitten*, Vol. 8, Immanuel Kant Werkausgabe, Frankfurt
— (1987), The Reason of Rules and the Rule of Reason, in: *Critica* 29, 43–86
— (2009), *Philosophy and Economics 1*, München
— (forthc.), ABC—Austria, Bloomington, Chicago. Political Economcy the Ostrom Way, in: *The Austrian and Bloomington Schools of Political Economy. Advances in Austrian Economics*, vol. 22, 15–47
Külpe, O. (1897), *Introduction to Philosophy a Handbook for Students of Psychology, Logic, Ethics, Æsthetics and General Philosophy*, New York
Leonard, R. (2010), *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*, Cambridge
McKenzie, R./G. Tullock (1978), *The New World of Economics: Exploration into the Human Experience*, Nueva York
Mises, L. von (1966[1949]), *Human Action*, Chicago
Morgenstern, O. (1928), *Wirtschaftsprognose: Eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*, Wien
Parsons, T. (1968), Utilitarianism. Sociological Thought, in: Sils, D./R. K. Merton (eds.), *International Encyclopedia of Social Sciences*, New York–London, 536–547
Rawls, J. (1951), Outline of a Decision Procedure for Ethics, in: *Philosophical Review* 60, 177–190

— (1971), *A Theory of Justice*, Oxford
Robbins, L. (1935), *An Essay on the Nature and Significance of Economic Science*, London
Schlick, M. (1986), *Die Probleme der Philosophie in ihrem Zusammenhang: Vorlesung aus dem Wintersemester 1933/34*, Frankfurt
Selten, R. (1999), Reply to K. Shepsle, in: Alt, J./M. Levi/E. Ostrom (eds.), *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science*, New York, 303–308
Spinoza, B. de (1951[1670]), *A Theologico-Political Treatise. A Political Treatise*, New York
Steinberg, J. (2013), Spinoza's Political Philosophy, in: *Stanford Philosophical Encyclopedia*
Strawson, P. F.(1962), Freedom and Resentment, in: *Proceedings of the British Academy* 48, 187–211
Tetlock, P. E. (2009), *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton
— (2015), *Superforecasting. The Art and Science of Prediction*, New York
Ypi, L. (2010), On the Confusion between Ideal and Non-ideal in Recent Debates on Global Justice, in: *Political Studies* 58, 536–555