

## Rawls' Theory of Justice ( Parts I and II ): Topics and Questions

0 Since we have only discussed the first three-hundred odd pages of **A Theory of Justice** in class, topics and questions you may wish to consider may extend beyond what we have covered. A few references to relevant later parts of the work will be provided, and the ( extensive ) index will yield many more.

Three primary Problemstellungen ( "problem-positings" ) tend to recur in discussions of **A Theory of Justice**, and ours have been no exception.

0.1 The first queries the epistemic/metaphysical/regulative complexity of Rawls's appeals to

0.1.1 **reflective equilibrium**, and

0.1.2 **the original position**.

0.2 The second focuses on some of the motivations and potential inadequacies of  
0.2.1 **maximin criteria** in general, and the **difference-principle** in particular, and

0.2.2 a complex of views Thomas Pogge calls Rawls' "**semiconsequentialism**", which tends to bracket from consideration "natural" ( as opposed to "social" ) "primary goods"

0.3 The third examines

0.3.1 Rawls' insistence on the **lexical priority** of his first principle of justice, at least "in reasonably favorable conditions", and

0.3.2 plausible **critiques** of that insistence -- as perturbations, perhaps, which may give rise to eventual alternative states of "reflective equilibrium" ( ? ).

In this handout, we will confine ourselves to some issues associated with the problems raised in 0.1 above. A later sheet will consider the questions raised in 0.2 and 0.3 in the light of Pogge's observations, and some separate arguments which have arisen in our own discussions in class.

\* \* \* \*

1 A central 'higher-order' **methodological problem** of Rawls' text is the elusive but extraordinarily suggestive **metatheoretic role** played by the "**original position**" and the "**parties**" therein, as they seek "**reflective equilibrium**".

In class, I suggested that there may something subtly self-referential about this role ( these roles ), and that this dialectical complexity may be witnessed by the ( 'transcendental' ? ) nature of his more general appeals to the idea(1) of "**reflective equilibrium**" ( cf. 20-21, 48-51, and passim ).

How -- I tried to ask -- should we **construe** this notion, in the end?

Is it in fact, as Rawls suggests, a **fixed point** for reflective inquiry? Or just a petitio of 'itself'?

Is it a **heuristic extrapolation** of a moral version of Aristotelean *theoria*, Or a metatheoretic counterpart of one of Kant's "**ideas of pure reason**"?

Is it only a **regulative ideal**, as Rawls sometimes seems to suggest?

Or a more ambitious ( and **constitutive**? ) "**Archimedean point**", a putative 'equilibrium'-state for forms of conceptual feedback which simply do not converge?

2 '**The state of reflective equilibrium** is to be achieved in the **original position**, of course, and the latter is thought-experimentally designed to facilitate it ( see below ). Rawls acknowledges **that there may well be no unicity** to such 'states', but urges us to argue heuristically ( or hypothetically ) **as if there were**.

He also acknowledges, very clearly and explicitly, that he has **carefully designed** the "constraints" imposed on the original position to mediate **the specific conclusions about justice** which then emerge ( as they should ) from ( his interpretation(s) of ) reflective equilibrium ( cf., e. g., 18-21, 141, 155, 579-80, 583-85 ).

He finally remarks, in the early section which introduces the original position, that "it should be impossible [ **in the original position** ] to tailor principles to the circumstances of one's own case" ( 18 ).

But is he **not**, then, engaged after all in a deeply-researched **metatheoretic**

counterpart of just such "tailoring"? Not to "one's own case", of course; but to a **collective** "case", of a distinctly more sympathetic sort, in which a society of like-minded "parties" engage in a spiritual exercise "tailored" to facilitate a carefully moderated form of quaker consensus?

3 Rawls grants explicitly at several points that he **does** essentially have to presuppose the unanimity mentioned above, or at least some moral-psychological convergence toward it, in the original position ( cf., e. g., 139 ), if he hopes to elicit from a concomitant 'reflective equilibrium' his carefully nuanced mixture of center-left social democratic principles, as outlined in **A Theory of Justice**.

He is also knowledgeable about the uses of metatheoretic arguments in other contexts ( cf., for example, his appreciative remarks about metamathematics on 51-52 ), and he is well-aware that certain methodological critiques naturally attach to such assumptions ( the basic one can be roughly paraphrased as an amiable heckle, perhaps from one from one of the "parties" themselves: "**Left-liberalism in/left-liberalism out( ?! )**" ). Accordingly, he offers a conscientiously nuanced palette of responses to such critiques. The two principal ones may be sketched roughly as follows.

The first, more pragmatic and concessive response is outlined at various points in the text, when Rawls himself queries the resulting theories adequacy and scope( cf. especially 18-21 and 577-82 ).

In such passages, he characterises the original position as an "expository device"; a thought-experimental state which we can enter "at any time, . . . so to speak, by following a certain procedure" ( 19 ); an "intuitive notion which suggests its own elaboration" ( 21 ), and enables us

. . . to see if the principles which would be chosen match our considered convictions of justice or extend them in an acceptable way" ( 19 ). . . . I doubt. . . that the principles of justice ( as I have defined them ) will be the preferred conception on anything resembling a complete list [of alternatives]. ( 581 )

The second response appears in other, more assertive passages -- the following, for example, in which Rawls insists that

acceptance of these principles is not conjectured as a psychological law or probability. . . . To be sure, the parties in the original position have a certain psychology, since various assumptions are made about their beliefs and interests. . . . But clearly arguments from such premises can fully deductive, as theories in politics and economics attest. **We should strive for a kind of moral geometry**, with all the rigor which this name connotes ( 121, emphasis mine ).

. . . justice as fairness is not at the mercy, so to speak, of existing wants and interests. **It sets up an Archimedean point for assessing the social system without invoking a priori considerations** ( 261, emphasis mine ).

Which is it?

4 Consider a ( highly conjectural ) **parallel**, between the following two assertions:

4.1 "A self-referential regulative ideal of collective '**rationality**' ( a fixed point of a search for 'itself' )  $\approx$  '**science**'; and

4.2 "A self-referential regulative ideal of collective '**reasonableness**' ( a fixed point of a search for 'itself' )  $\approx$  '**justice**'".

Might parallels such as **this** be the sort of thing Rawls has in mind?

5 Rawls' appeals to Kant suggest the following ironic juxtaposition, implicit in his assimilation of the original position to Kant's Reich der Zwecke.

5.1 In the **Grundlegung**, consideration of the "**realm of ends**" inevitably requires **counterfactual thought-experimentation**. Since this "realm" is **utterly noumenal**, such thought-experimentation remains **infeasible**, except as a limiting **regulative ideal** ( which transcends any actual experience ).

5.2 In A Theory of Justice, by contrast, consideration of the ambiguities of **actual experience** requires (counter-)counterfactual thought-experimentation. Since actual experience is **illimitably complex**, such thought-experimentation on the part of the "parties" remains **infeasible**, except as a **limiting framework of general theory** ( which **underdetermines** actual experience ).

With this and other contrasts in mind, I argued in class that the principal guiding motive of Kant's ethics was **not** really "**contractarianism**" ( contrary to Rawls' interpretation, sketched in 178-9, 251-7 and elsewhere ), but a monistic sort of "**intuitionism**". On the view Kant offers ( or so it seems to me ), the ultimate **unicity** of his "moral law within us" 'must' simply be **assumed** -- or perhaps less sympathetically ( more "transcendentally"? ): **begged** ).

Despite my belief that this is so, it seems **equally** clear to me that there **are**, indeed, many suggestive parallels between Kant's 'reasonable faith' and ( what might be described as ) Rawls' 'secularisation' thereof.

Now that we have all worked through some of the relevant texts, what do **you** think about the interrelations between Rawls' views and Kant's?  
6 The formulation of 5.2 above suggests ( or recalls ) other methodological query about the forms of thought-experimentation Rawls associates with the initial position. A recurrent need to **test** his "general" and "special conceptions of justice" against the more complex "constraints" of the 'real world' would be hard to dissociate from the apparent implications of several of his texts.

The remark from page 19, for instance, quoted above -- that we "can enter the original position at any time, so to speak, simply by following a certain procedure" ( compare Søren Kierkegaard's "movements of infinity" ) -- may well bear this interpretation.

More generally, the quasi-**dynamic** process which is supposed to eventuate in reflective equilibrium seems hard to **conceive**, without **revisions** in assumptions **initially** entertained behind the veil of ignorance -- revisions which would then carry over to "provisional" conclusions reached at the first of Rawls' four "stages", not later, "constitutional" and "legislative" ones.

Where, for example, would **revisions** in the "parties'" knowledge of "general facts" -- about "basic goods", "basic liberties" and "social and psychological theory", for example -- **come** from, if not from ( sufficiently "generalised" and anonymous ) **information**, which would have to (re)filter through the veil -- information, for instance, about "**the long-term prospects of the least favored extending over future generations**" ( 285 )?

If such 'hypothetical' ( ? ) information were needed to **reach** as well as **revise** reflective equilibrium, would this need not come into conflict with one of Rawls' most stringent "**constraints**" on the original position -- that the "parties" take '**everything**' ( ? ) into due consideration from the very start, since their decision will be **ultimate, binding and "final"** ( cf. 135ff and 176-177 )?

If reflective equilibrium, in short, is to be achieved by "reflection" on the implications of "fixed points" Rawls **acknowledges** are "provisional", **how** could the required revisions **come about**? For it seems that such revisions would have to mediate between one "original position" and **many** potential "social orders" extending through time, yet **all** these scenarios ( including, presumably, the "original position" itself ) are supposed in some sense to be "**closed systems**" ( cf., e. g., 8 ). Do the essentially sympathetic criticisms of Pogge and others not **bear witness**, in effect, to the inevitable '**openness**' and **plurality** of ( revised versions of ) Rawls' "constraints" -- as, indeed, on Rawls' own more concessive view of his "regulative" aims, they probably should?

Perhaps the very '**ideal**' of a social "**closed system**", hypothetical or '**real**', may -- like some other, even more venerable "**Archimedean points**" in the history of philosophy -- be a "**transcendent(al) illusion**"?

