The Common Ground – Self-organization

Marcello Di Bello - ASU - Spring 2023 - Week #12

The topic for today is self-organizing moral systems—specifically, under what conditions (if at all), people who initially disagree about questions of morality and politics can nevertheless converge to a shared set of moral ideals as they interact with one another.

We will read the work of Gerald Gaus, one of the main proponents of self-organizing moral systems.[1] Gaus positions himself in stark contrast with the contractualist tradition in political philosophy which culminates with John Ralws.

[1] See, in particular, Gaus (2017), Self-organizing moral systems: Beyond social contract theory, *Philosophy, Politics and Economics*, 17(2). See also Gaus (2018), The Complexity of a Diverse Moral Order, *The Georgetown Journal of Law & Puplic Policy*, 16(S)}.

Political liberalism recap

The hope of political liberalism, following Rawls's notion of an *overlapping consensus*, is that reasonable people will converge to a set of shared values and ideals that ground a conception of political justice despite the fact of reasonable pluralism. As Rawls puts it, an overlapping consensus is "affirmed by opposing religious, philosophical and moral doctrines likely to thrive generations."[2].

So, the overlapping consensus does not contain any specific conception of justice or the good life: it is political, not metaphysical, to use Rawls's famous phrase. But what does this conception of political justice grounded in an overlapping consensus look like?

> the conception I have elsewhere called 'justice as fairness' is a political conception of this kind. It can be seen as starting with the fundamental intuitive idea of political society as a fair system of social cooperation between citizens regarded as free and equal persons . . . The problem of justice is then understood as that of specifying the fair terms of social cooperation between citizens so conceived. The conjecture is that by working out such ideas, which I view as implicit in the public political culture, we can in due course arrive at widely acceptable principles of political justice.[3]

[2] John Rawls (1987), The Idea of an Overlapping Consensus, *Oxford Journal of Legal Studies*, 7(1), p. 1

[3] ibidem, p. 7

In short, this overlapping consensus contains two key tenets: (a) citizens are free and equal; and (b) a political society is a fair system of social cooperation. The problem of political justice, then, is the problem of specifying the fair terms of social cooperation between free and equal citizens.

Rawls does not say how this overlapping consensus comes about, however. He conjectures that, in a democratic liberal society such as ours, an overlapping consensus will emerge and stabilize over time. There might still be unreasonable citizens who hold views that are incompatible with the overlapping consensus—say, they think citizens are not equal, or they think that a political society should

not be a fair system of cooperation—but they will tend to reside on the fringes. Some philosophers in the Rawlsian tradition, such as Jonathan Quong, argue that the state can legitimately suppress unreasonable views when they pose an existential threat to the liberal state.[4]

[4] See chapter 10 of Quong, *Liberalism without Perfectionism*, Oxford University Press.

I/we believe we ought

Rawls' political philosophy falls in the so-called contractualist tradition. This tradition intends to solve a seemingly intractable social coordination problem.

Suppose you are I hold different conceptions of what is good. As we decide to act in light of what we each think we should do, our actions may come into conflict. Using Gaus's terminology, this is the approach (i) "I believe we ought" to social morality. But this approach cannot address the coordination problem because different people may hold conflicting beliefs about we ought to do, such as I believe we ought to φ while you believe we ought *not* to φ.[5]

[5] For example, I believe we ought to erect a fence around houses to protect people's private property from theft; you instead believe we ought to share land to grow food.

In order to avoid these conflicts, we should agree on shared rules of cooperation and then base our deliberations about what to do on these shared rules. What we need, then, is a conception of social morality (or a conception of political justice) that speaks to the questions of what (ii) "we believe we ought to do", not simply what (i) "I believe we ought to do". But how we do establish these shared rules of cooperation? What if—as is very likely—we do not actually believe the same thing about what we ought to do? The contractualist tradition—Gaus holds—has managed to only address the question of what (iii) "a theorist believes that we believe we ought to do", not so much the question of what (ii) "we believe we ought to do".

> a theory that acknowledges that we disagree about justice, yet need to coordinate, is certainly a great improvement upon simple "I believe we ought to" reasoning, as it seeks to confront at a basic level the fundamental moral insight that unless you and I concur about the demands of justice, our social relations will be deeply flawed from the perspective of justice itself. Yet, at the end of the day, it is a theory of what the theorist believes that we all believe what we all ought to do. That is, at the end of the day, it is one person's conviction about what we all believe we ought to do; and for the same reasons we disagree in our simple "I believe we ought" judgments, we disagree in our "I believe we believe we ought" judgments. (p. 7-8)[6]

[6] Page numbers (here and below) refer to the paper by Gaus, Self-organizing moral systems, following the pre-print version posted on the course website.

So, ultimately, while Rawls's overlapping consensus aims to address questions about what we believe we ought to do, it only answers the more parochial question of what Rawls (and those who think like him) believe we ought to do.

> social contract views are "top-down" (from the philosopher to us) theories of what a "bottom-up" (what we collectively would choose) morality might look like. (p. 11)

But is there an alternative? What does bottom-up morality look like?

Moral pluralism

Let there be a plurality of moral rules R_i's, more or less along the lines of Rawls's plurality of conceptions of the good.[7] Gauss formulates the question of a bottom-up social morality, as follows:

[7] Rawls talks about reasonable pluralism. How do Rawl's and Gaus's conceptions of moral pluralism (or diversity) differ from one another?

> Would agents who disagree in their (i) "I believe we ought" judgments of justice and (ii) judgments of the relative importance of reconciliation, converge on common rules, or would they each go their own way? Under what conditions might free individual moral reasoning replace the collectivist constructivism of the social contract? (p. 11-12)

This is an empirical question that cannot be theorized from the armchair. As apparent in the quotation, Gaus's moral pluralism is modeled along two key dimensions. They can be spelled out thusly:

(i) each agent assigns an *inherent* preference score, say between 0 and 10, to moral rules R_i's, where the score is inherent in the sense that it does not take into account whether other agents share the same assessment of the rule in question (denote this by $\mu_A(R_i)$, with A the agent in question and R_i the given moral rule);

(ii) each agent assigns a *relational* preference weight[8] between 0 and 1 to moral rules R_i's, where the weight depends on how many other people share the same assessment of the rule (denote this by $w_{B(n)}(R_i)$ where B is the agent in question, R_i the given moral rule, n the number of agents who agree with agent B about R_i).

[8] Gaus focuses on four paradigmatic relational weighting preferences, which he calls: quasi-Kantian; linear agents; moderately conditional agents; highly conditional cooperators; see Figure 4 on page 13.

The two dimensions reflect a key feature of moral pluralism, that is, some people care about whether others share their same preference for certain moral rules, while others might not. [9]

So, each agent A will assign an *overall utility* U_A to a given moral rule R_i, as follows:

$$U_A(R_i) = \mu_A(R_i) \times w_{A(n)}(R_i).$$

In other words, the utility is the inherent preference score μ assigned to rule R_i adjusted by the relational preference weight w. Whenever a rule R_1 has greater utility than another rule R_2 for the agent A—that is, $U_A(R_1) > U_A(R_2)$—agent A will make decisions in accordance with R_1 and not R_2.

[9] "In this analysis, then, an agent is concerned with both his own evaluations of the inherent justice of a rule (i.e., his "I believe we ought" conclusions) and reconciliation with the judgments of others ("we believe we ought"), and will ultimately make his decision based on his own view of the inherent justice of the rule given his evaluative standards and the weighted number of others who are acting on the rule." (p. 14)

Convergence

With a definition of moral pluralism in place, we can now ask whether there is any hope of bottom-up convergence. To answer this question, Gaus deploys computer simulations. He relies on different simulation models that mimic how individuals who hold different conceptions of the good—they prefer different moral rules—might end up converging or diverging as they interact with one another.

Gaus runs three different models in which 101 agents start by holding different preferences about two competing moral rules R_1 and R_2, some preferring the former and others preferring the latter.

Model I - fully random: People randomly assign scores μ between 0 and 10 to R_1 and R_2. Agents are randomly assigned one of the four weighting types. The starting point is an almost even split, where 51 people prefer R_2 and the rest prefer R_1. In this model, after a few iterations, everybody ends up preferring R_2.[10]

Model II - moderate polarity: What if society is initially polarized? Suppose some people strongly prefer R_1 (that is, $\mu(R_1)$ is high for them), while other people prefer R_2 (that is, $\mu(R_2)$ is high for them). So long as the initial split is 56/45 (that is, 56 people assign $\mu(R_1)$ high and 45 $\mu(R_2)$ low), there is convergence to R_1 after a few iterations.[11] However, if the initial split is more evenly balanced, say 52/49, there is no convergence even in the long run.

Model III - different reference groups: In principle, agents may have narrower reference groups than the entire society. That is, they might care about agreement with others so long as they belong to their reference groups. So the relational weighting preference can be relativized to reference groups smaller than the entire society. What happens when reference groups are allowed to vary? Convergence is still possible under certain conditions.[12]

These are basic models and one might wonder to what extent they are realistic. Gaus's aim, however, was to sketch an alternative to the approach by the social contract tradition (including Rawls's project) which handles pluralism by seeking an underlying homogeneity (an "overlapping consensus"). Gaus's project is radically different:

> The guiding idea is to model morally autonomous diverse agents making choices in the context of each other's choices, seeing what dynamics lead to a shared rule that all endorse, and when different groups will go their own way. The motto of this project is that morality is best understood as a bottom-up affair. "The moral law is not imposed from above or derived from well-reasoned principles" but arises from the values of individuals and their distinctive searches for integrity and reconciliation in their social-moral lives. (p. 26)[13]

[10] See Figure 5, p. 17. The omission of highly conditional cooperators from the simulation slowed down the process of convergence, but did not make it impossible; see Figure 6, p. 18

[11] As before, the omission of highly conditional cooperators from the simulation slowed down the process of convergence, but did not make it impossible; see Figure 7, p. 20

[12] See Figure 9, p. 23.

[13] One worry here is that the content of this bottom-up morality is unknown, but this seems to be precisely Gaus's point. No theorist can legislate in advance about the principles of this bottom-up morality.