Working Paper

Working Paper 00 (2023) 1–4

Socially optimal ideology

Richard Povey

Hertford College, Catte Street, Oxford, OX1 3BW.

September 25, 2023

Abstract

A simple sequential game is used to show that the level of social welfare achievable in equilibrium is non-monotonic in the level of divergence in agents' perceptions of the optimal state of the world. In particular, although zero divergence yields globally maximised social welfare, some finite levels of divergence are Pareto-dominated by higher levels. Hence a world in which agents are more agreed about the optimal state is not necessarily one in which agents are better off.

Keywords:
ideology, welfare, efficiency
JEL:

1. *Introduction*

> *If men were angels, no government would be necessary. If angels were to govern men, neither external nor internal controls on government would be necessary.* *(Madison, 1788)*

> *If even a pebble lies where I want it to lie, it cannot, except by a coincidence, be where you want it to lie.* *(Lewis, 1942)*

Ideological disagreement, in the sense that individual agents diverge in their perception of the optimal state of an external world, is an essential prerequisite for the existence of any economic resource allocation problem. It also transcends the distinction between self-interest and altruism. Economists are usually content to merely assume that this is the way that the world (and agents' preferences about it) works without asking the meta-question of *why* this should be so. However, given increasing empirical evidence that "moral preferences" evolve culturally and differ between societies, a functionalist explanation of this phemonenon would seem to be desirable.

This paper presents a simple sequential game theoretic model in which the efficiency of the equilibrium outcome is non-monotonic in the level of ideological diversity. The intuition for this result is that greater ideological disagreement is a "double-edged sword" in that it creates stronger initial incentives for misbehaviour, whilst also enabling more effective punishment after such transgressions occur. Hence, as well as the obvious global social optimum where there is complete ideological agreement (and hence no economic problem), there can exist local optima where a small *increase or decrease* in ideological diversity is welfare-reducing.

☆Email: *richard.povey@hertford.ox.ac.uk*

2. The Model

Consider a continuum of different types of individual agent, each of which has an optimal state of the world x, which must be a real number distributed according to the p.d.f. $g(x)$ with support in $[0, s]$ so that $\int_0^s g(x)dx = 1$. Let $\mu = \int_0^s xg(x)dx$ denote the expected value of this distribution and $\sigma^2 = \int_0^s x^2 g(x)dx - \mu^2$ denote the variance. We assume a symmetric distribution so that $\mu = {}^1\!/_2 s$. The optimal state for a given agent is assumed to be private information for that agent. The game consists of an infinite number of periods. All agents are assumed to share the same discount factor $0 < \delta < 1$. In period t, a distinct agent (with optimal state x_t) is selected from the continuum and is able to set the *current* state of the world z_t. The state of the world set (z_t) is publicly observed by all agents at time t. Hence all agents selected to make a move from time $t + 1$ onwards can condition their strategy upon z_t. Agents are assumed to have quadratic utility so that the present discounted value of utility looking forward from time t for type x_t will be:

$$u_t = -\sum_{i=0}^{\infty} \left[\delta^i (x_t - z_{t+i})^2\right] \tag{1}$$

3. Anarchic Equilibrium

The simplest conceivable subgame perfect Nash equilibrium for the game would be for each individual to set their optimal state of the world so that $\forall_t : z_t = x_t$. No threats conditioned on observed past behaviour would need to be made in this case in order to sustain the equilibrium. Social welfare (utility for a representative agent i who is behind a "veil of ignorance", not yet knowing the value of x_i) in a single period t would then be:

$$w = E_{x_i,x_t}\left[-(x_i - x_t)^2\right] = -\int_0^s \int_0^s (x - y)^2 g(x)g(y)dx\,dy \tag{2}$$

Since x_i and x_t have the same mean μ and variance σ^2 and are independently distributed (so that $Cov[x_i, x_t] = 0$), we then have:

$$w = -E_{x_i,x_t}\left[(x_i - x_t)^2\right] = -\left(Var_{x_i,x_t}\left[(x_i - x_t)\right] + \left(E_{x_i,x_t}\left[(x_i - x_t)\right]\right)^2\right) = -2\sigma^2 \tag{3}$$

4. Punishment Equilibrium

Due to the increasing marginal disutility generated by quadratic preferences as the difference between z_{t+i} and x_t increases, it would generate higher social welfare if all agents were to set $z_t = \mu$ in every period t. Indeed, it can easily be shown that setting $\forall_t : z_t = \mu$ is the socially optimal equilibrium.[1] Social welfare per period will then be:

$$w = E_{x_i}\left[-(x_i - \mu)^2\right] = -Var_{x_i}[x_i] = -\sigma^2 \tag{4}$$

In order to sustain such an equilibrium, clearly some form of credible punishment must be used in order to deter agent t from setting $z_t = x_t$ instead of $z_t = \mu$. The most obvious punishment scheme would be to revert to the anarchic equilibrium in all future periods if any agent deviates in period t by setting $z_t \neq \mu$. For agent t, the net gain in utility from co-operating by setting $z_t = \mu$ will be:

[1] This can be done simply by differentiating the expected per period utility of the representative agent with respect to z and setting equal to 0 for the first order condition:

$$w(z) = E_{x_i}\left[-(x_i - z)^2\right] = -\int_0^s (x - z)^2 g(x)dx$$

$$\frac{dw}{dz} = \int_0^s 2(x - z)dx = 0$$

$$\implies \int_0^s xg(x)dx = z^* \int_0^s g(x)dx \qquad \implies z^* = \mu$$

$$\lambda(x_t) = \frac{\delta}{1-\delta}\left(\int_0^s (x_t - y)^2 g(y)dy - (x_t - \mu)^2\right) - (x_t - \mu)^2 \tag{5}$$

Differentiating with respect to x_t, we get:

$$\frac{d\lambda}{dx_t} = 2\left(\frac{\delta}{1-\delta}\left(\int_0^s (x_t - y)g(y)dy - (x_t - \mu)\right) - (x_t - \mu)\right) = -2(x_t - \mu) \tag{6}$$

The second derivative is:

$$\frac{d^2\lambda}{dx_t{}^2} = -2x_t \tag{7}$$

Since $\frac{d^2\lambda}{dx_t{}^2} \leq 0$, $\lambda(x_t)$ is a weakly concave function with $\frac{d\lambda}{dx_t}\Big|_0 = 2\mu > 0$, and so a sufficient condition to have $\forall_{0\leq x_t\leq s} : \lambda(x_t) \geq 0$ is that $\lambda(0) \geq 0$ and $\lambda(s) \geq 0$:

$$\lambda(0) \geq 0 \Longrightarrow \frac{\delta}{1-\delta}\left(\int_0^s (y)^2 g(y)dy - \mu^2\right) - \mu^2 \geq 0 \Longrightarrow \sigma^2 \geq \left(\frac{1-\delta}{\delta}\right)\frac{1}{4}s^2 \tag{8}$$

$$\begin{aligned}
\lambda(s) \geq 0 &\Longrightarrow \frac{\delta}{1-\delta}\left(\int_0^s (s - y)^2 g(y)dy - (s - \mu)^2\right) - (s - \mu)^2 \geq 0 \\
&\Longrightarrow \frac{\delta}{1-\delta}\left(\int_0^s (2\mu - y)^2 g(y)dy - (2\mu - \mu)^2\right) - (2\mu - \mu)^2 \geq 0 \\
&\Longrightarrow \frac{\delta}{1-\delta}\left(4\mu^2\int_0^s g(y)dy - 4\mu\int_0^s yg(y)dy + \int_0^s y^2 g(y)dy - \mu^2\right) - \mu^2 \geq 0 \\
&\Longrightarrow \frac{\delta}{1-\delta}\left(\sigma^2 + \mu^2 - \mu^2\right) - \mu^2 \geq 0 \Longrightarrow \sigma^2 \geq \left(\frac{1-\delta}{\delta}\right)\frac{1}{4}s^2
\end{aligned} \tag{9}$$

We can see that (8) and (9) reduce to the same condition.

The highest possible variance for the distribution of x_t would be a Bernoulli distribution with a probability of $1/2$ for $x_t = 0$ and $1/2$ for $x_t = s$, yielding a variance of $1/2 \times (s - 1/2 s)^2 + 1/2 \times (0 - 1/2 s)^2 = 1/4 s^2$. This means that the punishment equilibrium can only exist if $1/4 s^2 \geq \frac{1-\delta}{\delta}\frac{1}{4}s^2 \Longrightarrow \delta \geq 1/2$.

5. *The Implied Social Welfare Function*

Using the standard deviation of x_t (σ) as our measure of ideological diversity, social welfare as a function of this will be:

$$w(\sigma) = \left\{ \begin{array}{cccc} -2\sigma^2 & if & \sigma^2 < \frac{1}{4}s^2\left(\frac{1-\delta}{\delta}\right) \\ -\sigma^2 & if & \sigma^2 \geq \frac{1}{4}s^2\left(\frac{1-\delta}{\delta}\right) \end{array} \right\} \tag{10}$$

Figure 1 graphs the implied social welfare function for the case where $s = 2$ and $\delta = 3/4$. Crucially, we can see that as well as the global maximum at $\sigma = 0$, there is a discontinuity and thus a local maximum where $\sigma = \frac{1}{2}s\sqrt{\frac{1-\delta}{\delta}} = \frac{1}{\sqrt{3}}$. If σ falls below this value then it may be possible to achieve a Pareto improvement in equilibrium by *increasing* σ.

It is also instructive to examine the critical value of σ below which an increase in σ cannot achieve a Pareto improvement. This is where $-\sigma^2 = -2\frac{1}{4}s^2\frac{1-\delta}{\delta} \Longrightarrow \sigma = \frac{1}{\sqrt{2}}s\sqrt{\frac{1-\delta}{\delta}}$. So, provided that $\delta \geq 1/2$ and $\sigma \in \left(\frac{1}{2}s\sqrt{\frac{1-\delta}{\delta}}, \frac{1}{\sqrt{2}}s\sqrt{\frac{1-\delta}{\delta}}\right)$ then a Pareto-improvement can be achieved through an increase in σ as well as through a decrease.

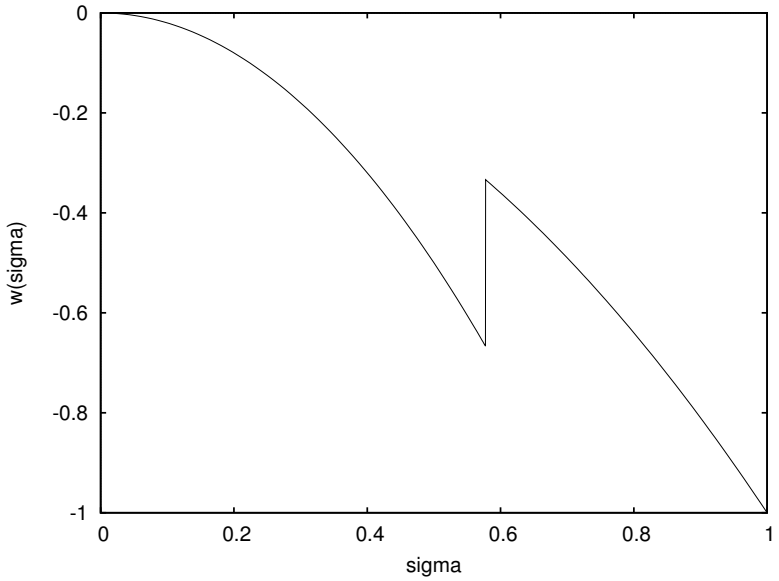

Figure 1. Social welfare as a function of σ with $s = 2$ and $\delta = 3/4$.

6. *Conclusion*

This paper has shown that, counter to everyday intuition, the sequential nature of interactions means that a society in which agents exhibit greater agreement about the optimal state of an external world is not necessarily one in which agents are better off. In fact, it will in general be possible to make Pareto improvements by increasing the level of ideological disagreement if the current level of divergence falls within a critical range.

References

Lewis, C.S., 1942. The problem of pain. London. URL: `http://nla.gov.au/nla.cat-vn1050470`.
Madison, J., 1788. The Federalist Papers. Yale University Press. URL: `http://www.jstor.org/stable/j.ctt5vm398`.