The Evolution of Monetary Equilibrium*

Thomas W. L. Norman

Magdalen College, Oxford

December 16, 2019

Abstract

The Hahn problem is that, even if a "monetary" equilibrium with valued fiat money exists in general equilibrium, a "nonmonetary" equilibrium with a zero price on money generally also exists; why should we expect the former over the latter? Here, I consider the preferences that will survive repeated trading in an exchange economy where agents compete in biological fitness. With unobservable preferences and positive assortativity in matching, evolutionarily stable preferences implement the competitive equilibrium that maximizes the sum of agents' fitnesses. In a standard Bewley–Townsend model, this implies selection of the monetary over the nonmonetary equilibrium, and also implies the survival of agents with "money in the utility function". *Journal of Economic Literature* Classification: C73, D5, E4

Key Words: money; Hahn problem; preference evolution; evolutionary stability; money in the utility function.

> Trust in fiat money is only a recent development, and even today such faith is hardly universal. (Cass and Shell 1980)

*Email: thomas.norman@magd.ox.ac.uk.

1 Introduction

The value of fiat money in general equilibrium poses some well-known problems: "There is always the possibility (in some models, the necessity) of the equilibrium price of fiat money to be nil" (Starr 1989, p. 292), demonetizing the model. This is the Hahn problem (1973)—to articulate general-equilibrium models where money is essential.[1] It "could be that all households would be better off if the price of money were positive. . . but this does not imply that zero price is a disequilibrium price" (Cass and Shell 1980). Hence, conditions for the existence and selection of a monetary equilibrium are of paramount interest.

There are of course some well-known responses to this problem in monetary theory. Some approaches assume an additional value of money: "money in the utility function" (MIU); the "cash-in-advance constraint"; and models where money can be used in payments to the government (Obstfeld and Rogoff 1983; Ritter 1995; Aiyagari and Wallace 1997; Li and Wright 1998). There are others that dispense with market completeness to give money value as a store of wealth for overlapping generations (Samuelson 1958; Wallace 1980) or insurance (Bewley 1980; Townsend 1980). And there are others still that dispense with the Arrow–Debreu frictionless, timeless centralized market to give money value as a matching lubricant (Kiyotaki and Wright 1989, 1993; Shi 1995; Trejos and Wright 1995), or because it is expected to have value in a temporary equilibrium (Grandmont 1974). The Fiscal Theory of the Price Level (Leeper 1991; Woodford 1995) pins down the value of money as a claim on the government's balance sheet, Kocherlakota (1998) has money play an essential role in recording transactions, and Wallace and Zhu (2004) justify fiat money as the limit of a commodity-money system (see also Araujo and Guimaraes 2014).

Thus, whilst the Hahn problem can hardly be said to be unresolved, perhaps the most widely used response is the MIU approach, albeit to the disapproval of a sizeable number of economists, who regard it as a reduced form at best and *ad hoc* at worst. But why is there such adamance that fiat money should not appear in the utility function? Wallace (1996), for instance, offers the dictum that "money should not be a primitive in monetary theory". For Lagos, Rocheteau, and Wright (2017, p.

[1] This is distinct from the "modified Hahn problem" (Hellwig 1993) or "rate-of-return dominance puzzle" of why fiat money should survive in the presence of substitutable interest-bearing assets, which this paper does not address. Wright (1995) does analyse the question of which of multiple goods might serve as money from an evolutionary perespective.

332), meanwhile, anything (such as MIU) that deviates from the discipline of deriving the value of assets endogenously "is obviously subject to the Lucas critique". And indeed, it seems compelling that money should not feature in the utility function *by assumption*. But that does not mean that it should not feature in the utility function *at all*; it merely calls for a theory that explains money's appearance in the utility function endogenously.[2] This leads naturally to the theory of preference evolution.

Preference evolution views preferences not as fixed and exogenous, but rather as subject to selection according to their success in an underlying biological contest. One particularly prominent approach in this field has been that of "indirect" preference evolution (Güth and Yaari 1992), whereby utility is determined according to the biological "fitness" that it brings in some underlying game. When such utility is observable, it carries a commitment power, which often favors efficient outcomes enforced by non-fitness-based preferences (Dekel, Ely, and Yilankaya 2007). However, when preferences are unobservable and matching is nonassortative, these results break down in favor of Nash outcomes enforced by players maximizing fitness (Ely and Yilankaya 2001). Nevertheless, if matching is assortative, Alger and Weibull (2013) demonstrate that stable unobservable preferences still depart from biological fitness. In particular, in such circumstances, evolution favors a utility function that is a convex combination of standard self-interest and "Kantian" preferences that evaluate a choice on the assumption that others will also choose it.

This Kantian motivation represents an interesting force from the perspective of the Hahn problem, for it suggests a mechanism for coordinating on monetary equilibrium if there are collective gains to doing so. Indeed, the theory of preference evolution offers a natural expression of the idea that fiat money is intrinsically worthless—that it should yield no *biological fitness* directly. This seems unarguably to be the case; pieces of paper give me no reproductive advantage in and of themselves. However, it is perfectly plausible that evolution might lead me to derive *utility* from such pieces of paper, if this leads me to behave in a way that is to my reproductive advantage. Moreover, in so doing, it might *select* a monetary equilibrium over a nonmonetary one, thus addressing the Hahn problem.[3]

This is the nature of the results that I establish in this paper. To begin with,

[2] MIU also has the attraction that it is the only approach to a monetary equilibrium that can leave the centralized, timeless competitive approach intact.

[3] Admittedly, the set of monetary equilibria may still be large (Matsuyama 1991).

I specify a general two-agent exchange economy that allows for the possibility of money, and then offer a Bewley–Townsend example that has both a monetary and a nonmonetary equilibrium, as per the Hahn problem.[4] I go on in Section 3 to specify a population game amongst individuals matched to trade as in Section 2. This game fits the model of Alger and Weibull (2013), and hence implies the selection of preferences that place some weight on what would happen if all individuals played the same way. As a result, evolutionarily stable preferences under positive assortativity implement the competitive equilibrium from Section 2 that maximizes the sum of individual fitnesses. This equilibrium need not be Pareto-optimal (in fitnesses), but there can exist no other equilibrium Pareto-superior to it. In the Bewley–Townsend example, the monetary equilibrium is selected; by contrast, a type that is programmed to play the nonmonetary equilibrium is evolutionarily unstable.

2 The Hahn Problem

Suppose that there are two agents $i = A, B$ who trade a single consumption good and fiat money in an exchange economy through discrete time $t \in \mathscr{T} \subseteq \mathbb{N}$. Agent A is endowed with $\omega_t^A \in [0, C]$ units of the consumption good in period t, and $m_{-1}^A \in [0, M]$ units of unbacked, intrinsically useless money prior to trading; agent B is endowed with $C - \omega_t^A$ units of the consumption good in period t, and $M - m_{-1}^A$ units of money prior to trading. Each agent must determine his *plan* $x = (c^i, m^i)$, consisting of a *consumption plan* $c^i \equiv (c_0^i, c_1^i, \ldots) \in [0, C]^{\mathscr{T}}$ and a *money plan* $m^i \equiv (m_0^i, m_1^i, \ldots) \in [0, M]^{\mathscr{T}}$, given the common *payoff* function $v(c_t^i)$, where $p_t \in \mathbb{R}_+ \cup \{\infty\} \equiv \mathscr{P}$ is the (relative) price of consumption in period t. Thus, money does not affect agents' payoffs, and may also have no value, since p_t may take the value ∞ in the one-point compactification of the positive real line, $\mathbb{R}_+ \cup \{\infty\}$.[5] Call this economy $\mathscr{E}$, and suppose that it is *regular* in the sense of Debreu (1970).

Letting $p \in \mathscr{P}^{\mathbb{N}}$ be the *price path* through time, *competitive equilibrium* requires

[4] Note that the results are not specific to bilateral trade, but rather could be extended to a multi-agent aggregative setting using Alger and Weibull (2016).

[5] Such a compactification of the price set is of course common when dealing with relative prices in general equilibrium, for instance in the classical proof of equilibrium existence.

that markets clear when each agent i solves the problem[6]

$$\begin{aligned}
&\max_{c^i, m^i} \sum_{t \in \mathscr{T}} \beta^t v(x_t), \\
&\text{s.t.} \quad p_t c_t^i + m_t^i \le p_t \omega_t^i + m_{t-1}^i, \qquad \forall t \in \mathscr{T} \\
&\qquad \forall t, i \quad c_t^i \in [0, C], \; m_t^i \in [0, M].
\end{aligned} \tag{1}$$

Thus, agents' endowments are only storable through money. The Lagrangian for this problem is

$$\mathscr{L} = \sum_{t \in \mathscr{T}} \beta^t \left(v(x_t) + \lambda_t^i \left[p_t \omega_t^i + m_{t-1}^i - p_t c_t^i - m_t^i \right] \right) \tag{2}$$

subject to $c_t^i \ge 0$, $m_t^i \ge 0$, where $\{\lambda_t^i\}_{t \in \mathscr{T}}$ is a sequence of nonnegative Lagrange multipliers. Let $\hat{x} = (\hat{c}^i, \widehat{m}^i)$ denote the optimal plan of an agent. A competitive equilibrium $(\overline{x}^A, \overline{x}^B, \overline{p})$ will be described as *maximal* if it maximizes $\sum_{t \in \mathscr{T}} \beta^t v(\overline{x}_t^A) + \sum_{t \in \mathscr{T}} \beta^t v(\overline{x}_t^B)$ in the set of competitive equilibria. Whilst maximality is neither necessary nor sufficient for a competitive equilibrium to be a Pareto-optimal allocation, there can exist no competitive equilibrium that is Pareto-superior to a maximal competitive equilibrium.

Example Consider Sargent's (1987, §6.6) infinite-horizon example of the incomplete-markets model of Bewley (1980) and Townsend (1980), where all private loan markets are closed. Agent A is endowed with $\omega^A \equiv \{\omega_1^A, \omega_2^A, \omega_3^A, \omega_4^A, \ldots\} = \{1, 0, 1, 0, \ldots\}$ units of the consumption good and $m_{-1}^A = 0$ units of unbacked money prior to trading; agent B is endowed with $\omega^B \equiv \{\omega_1^B, \omega_2^B, \omega_3^B, \omega_4^B, \ldots\} = \{0, 1, 0, 1, \ldots\}$ units of the consumption good and $m_{-1}^B = M$ units of unbacked money prior to trading. The first-order necessary conditions for a solution to (1) are

$$\begin{aligned}
\beta^t [v'(x_t^i) - \lambda_t^i p_t] &\le 0, \qquad = 0 \text{ if } c_t^i > 0, \qquad t = 0, 1, \ldots \\
-\beta^t \lambda_t^i + \beta^{t+1} \lambda_{t+1}^i &\le 0, \qquad = 0 \text{ if } m_t^i > 0, \qquad t = 0, 1, \ldots \\
\lim_{T \to \infty} -\beta^T \lambda_T^i &\le 0, \qquad \lim_{T \to \infty} \beta^T \lambda_T^i m_T^i = 0.
\end{aligned} \tag{3}$$

[6] This set-up could straightforwardly be extended to incorporate uncertainty and hence rational-expectations equilibrium in general.

The first two equations are equivalent to

$$v'(x_t^i) \leq \lambda_t^i p_t, \qquad = \text{ if } c_t^i > 0, \tag{4}$$

$$\frac{\lambda_{t+1}^i}{\lambda_t^i} \leq \frac{1}{\beta}, \qquad = \text{ if } m_t^i > 0. \tag{5}$$

Sargent (1987, §6.6) shows that there exists a competitive equilibrium of this economy with

$$c_t^A = \begin{cases} c^*, & t \text{ even} \\ c^{**}, & t \text{ odd} \end{cases} \qquad c_t^B = \begin{cases} c^{**}, & t \text{ even} \\ c^*, & t \text{ odd} \end{cases}$$

$$m_t^A = \begin{cases} c^{**}p, & t \text{ even} \\ 0, & t \text{ odd} \end{cases} \qquad m_t^B = \begin{cases} 0, & t \text{ even} \\ c^{**}p, & t \text{ odd.} \end{cases}$$

$c^* + c^{**} = 1$ and $p_t = p = M/c^{**}$ for all t. In this equilibrium, unbacked currency is valued and the agents pass money back and forth in an effort to smooth their consumption paths. However, note that (3), (4) and (5) are also satisfied by $p_t = \infty$, $m_t^i = 0$ for all t—i.e. there is another equilibrium where money has no value and is not demanded. Whilst this model allows a monetary equilibrium then, it also has an autarkic nonmonetary equilibrium, and it is not clear why we should select the former over the latter. Note for now though that the monetary equilibrium is not optimal (which would require a diminishing stock of money, as described in Sargent 1987, §6.7), but it is maximal. ■

3 *Homo Monetarius*

Suppose now then that an individual is matched to trade with an agent selected exogenously at random (and possibly assortatively) from a larger population of individuals. In this section, I define a game of this nature that fits within the framework of Alger and Weibull (2013): supposing that a certain "resident" preference type has become the predominant one in the population, do there exist other "mutant" preference types that can invade the population, by inducing equilibrium plans under which mutants' average fitness is at least as high as that of residents? Preference types that withstand such invasion by all other preference types are said to be evolutionarily stable. I show that, under assortative matching, evolutionary stability selects a maximal competitive equilibrium in general, and the monetary equilibrium of the Bewley–Townsend model in particular. The evolutionary framework within which I work is symmetric, and hence I symmetrize the game by having the two trading agents assume the roles (A, B) or (B, A) with equal probability (as per Selten 1980, Alger and Weibull 2013, §6.1).

Endow the space $\mathscr{P}^{\mathbb{N}}$ of price paths with the product topology (over the countable number of time periods), and do likewise for each of the spaces $[0, C]^{\mathbb{N}}$ and $[0, M]^{\mathbb{N}}$ of consumption and money plans. Suppose then that each individual has as a common strategy set X the topological vector space of continuous mappings from $\{A, B\} \times \mathscr{P}^{\mathbb{N}}$ into $[0, C]^{\mathbb{N}} \times [0, M]^{\mathbb{N}}$, endowed with the compact-open topology (which, given the metrizability of the codomain, coincides here with the topology of uniform convergence on compact sets). Since the domain of this function space is compact, X is equicontinuous, and since the codomain is also compact, X is pointwise relatively compact; it follows that X is itself then compact in the topology of uniform convergence on compact sets by the Arzelá–Ascoli theorem. Moreover, if any convex combination with $\{\infty\}$ is itself $\{\infty\}$, both the domain and the codomain of a strategy is convex, and hence so is the strategy space.

A typical strategy $x = (x^i)_{i \in \{A,B\}}$ specifies the individual's consumption and money plans (c^i, m^i), as a continuous function of the price path p, given the individual's role i. Given a pair (x, y) of strategies, each individual has a symmetric *fitness* function

$$\pi(x, y) = \frac{1}{2} \sum_{i \in \{A,B\}} \sum_{t \in \mathscr{T}} \beta^t v(x_t^i(\hat{p}_t(y))),$$

where $\hat{p}(y)$ is a market-clearing price path under the plans (y^A, y^B). Notice that fitness $\pi(x, y)$ is the expected value $V(x, \hat{p}(y))$ of the agent's discounted payoffs from the plan x along the price path $\hat{p}(y)$. Moreover, in common with the economy $\mathscr{E}$, the agents are *price-takers*; they are not able to manipulate the equilibrium price, which is determined from the opposing agent's plan y.[7]

Remark Is it reasonable for $\hat{p}$ to depend on y and not on x? Price-taking requires

[7] For the issues arising absent the price-taking assumption, see the rich literature exploring the implementation of the outcomes of competitive equilibria (Hurwicz 1979; Schmeidler 1980; Postlewaite and Wettstein 1983; Groves and Ledyard 1987; Hurwicz, Maskin, and Postlewaite 1995; Hart and Mas-Colell 2015) and expectations equilibria (Palfrey and Srivastava 1987, 1989; Blume and Easley 1990; Wettstein 1990) in games.

that the price be independent of x, but why then is it dependent on y? Because y captures the plans of the *rest* of the market, which here is just one other agent but more generally could be an arbitrary number of agents; and whilst one agent's plan becomes insignificant as the number of agents grows, the plans of all other agents do not. Hence, this formulation of $\hat{p}(y)$ is analogous to ruling out strategic play in a two-person exchange economy, which is the case of interest here. Moreover, as mentioned earlier, the model can straightforwardly be extended to a multi-agent aggregative setting using Alger and Weibull (2016). ∎

By virtue of $\mathscr{E}$'s regularity, a straightforward application of the Implicit Correspondence Theorem (see Mas-Colell 1985, Theorem J.3.3) yields that the price path $\hat{p}$ is a continuous function of y, and hence π is a continuous function of (x, y) (since $x(p)$ is continuous in both x and p).

Whilst biological fitness is thus determined by the trading outcome of $\mathscr{E}$, the individuals' preferences—and hence choices—are determined by evolutionary selection, according to their fitness in the resulting trades. This is the "indirect" evolutionary approach of Güth and Yaari (1992), but since preferences are assumed to be unobservable, the framework I use is that of Alger and Weibull (2013). Specifically, each individual has a private *type* $\theta \in \Theta$. It will be sufficient here to consider populations with two types present, as captured in the population *state* $s = (\theta, \tau, \varepsilon)$, where $\theta, \tau \in \Theta$ are the two types and $\varepsilon \in (0, 1)$ is the population share of type τ. If ε is small, θ is referred to as the *resident* type and τ the *mutant* type. In a given state $s \in S \equiv \Theta^2 \times (0, 1)$, let $\Pr[\tau|\theta, \varepsilon]$ be the probability that a given individual of type θ is matched with an individual of type τ.

An individual's type θ defines a *utility* function $u_\theta : X^2 \to \mathbb{R}$, which completes the description of a standard Bayesian game $\mathscr{G}$. No particular relation between u_θ and π is assumed, but u_θ is assumed to be continuous. For each state $s \in S$, any strategy $x \in X$ used by type θ and any strategy $y \in X$ used by type τ, the resulting average fitness to each type is

$$\Pi_\theta(x, y, \varepsilon) = \Pr[\theta|\theta, \varepsilon] \cdot \pi(x, x) + \Pr[\tau|\theta, \varepsilon] \cdot \pi(x, y),$$
$$\Pi_\tau(x, y, \varepsilon) = \Pr[\theta|\theta, \varepsilon] \cdot \pi(y, x) + \Pr[\tau|\theta, \varepsilon] \cdot \pi(y, y).$$

In any $s \in S$, meanwhile, a strategy pair $(x^*, y^*) \in X^2$ is a *Bayesian Nash equilibrium*

(BNE) of $\mathscr{G}$ if

$$\begin{cases} x^* \in \arg\max_{x \in X} \Pr[\theta|\theta,\varepsilon] \cdot u_\theta(x, x^*) + \Pr[\tau|\theta,\varepsilon] \cdot u_\theta(x, y^*), \\ y^* \in \arg\max_{y \in X} \Pr[\theta|\theta,\varepsilon] \cdot u_\tau(y, x^*) + \Pr[\tau|\theta,\varepsilon] \cdot u_\tau(y, y^*). \end{cases} \tag{6}$$

Alger and Weibull (2013) define evolutionary stability of types on the assumption that the resulting fitnesses are determined by this equilibrium set: a type $\theta \in \Theta$ is *evolutionarily stable against a type* $\tau \in \Theta$ if there exists an $\bar{\varepsilon} > 0$ such that $\Pi_\theta(x^*, y^*, \varepsilon) > \Pi_\tau(x^*, y^*, \varepsilon)$ in all BNE (x^*, y^*) in all states $s = (\theta, \tau, \varepsilon)$ with $\varepsilon \in (0, \bar{\varepsilon})$. A type θ is then *evolutionarily stable* if it is evolutionarily stable against all types $\tau \neq \theta$ in Θ. Evolutionary stability is a static equilibrium concept: the model focuses on Bayesian Nash equilibrium play during the whole lifetime of one given "generation" where a share ε of mutants is present in the population, and stability is checked by analyzing the case where ε tends to 0. Although there is time in the model, an individual's fitness is determined by aggregating over the consumption payoffs obtained over the totality of periods for which the interaction lasts, i.e. evolutionary success does not fluctuate over the course of the interaction. Evolution is thus occurring over a longer timescale than that of trade, consistent with the literature on "indirect" preference evolution with unobservable types.[8]

Let $B^{\mathrm{NE}}(s) \subseteq X^2$ denote the set of Bayesian Nash equilibria in population state $s = (\theta, \tau, \varepsilon)$—i.e. all solutions (x^*, y^*) of (6). Given types θ and τ, Alger and Weibull (2013) define an equilibrium correspondence $B^{\mathrm{NE}}(\theta, \tau, \cdot) : [0, 1) \rightrightarrows X^2$ that maps a mutant population share ε to its associated set of equilibria. Moreover, for each type $\theta \in \Theta$, they let $\beta_\theta : X \rightrightarrows X$ denote the best-reply correspondence,

$$\beta_\theta(y) = \arg\max_{x \in X} u_\theta(x, y) \quad \forall y \in X,$$

and $X_\theta \subseteq X$ the set of fixed points under β_θ,

$$X_\theta = \{x \in X : x \in \beta_\theta(x)\}.$$

Given $\theta \in \Theta$, Θ_θ is then the set of types τ that, when they constitute a small mutation amongst a population of resident type-θ individuals, play the same way as (and are

[8]Whilst the timescale of trade may be infinite in $\mathscr{E}$, the standard hazard-rate interpretation of the discount factor in that case will guarantee that each matched trading session concludes in finite time.

hence behaviorally indistinguishable from) those residents:

$$\Theta_\theta = \{\tau \in \Theta : \exists x \in X_\theta \text{ such that } (x, x) \in B^{\text{NE}}(\theta, \tau, 0)\}.$$

In the traditional case of uniform random matching, $\Pr[\tau|\theta,\varepsilon] = \Pr[\tau|\tau,\varepsilon] = \varepsilon$ for all $\varepsilon \in (0,1)$. Following Bergstrom (2003), Alger and Weibull (2013) allow a more general interaction technology, under which assortative matching is possible. Given $\theta, \tau \in \Theta$, $s = (\theta, \tau, \varepsilon)$, $\varepsilon \in (0,1)$,

$$\begin{aligned}
&\phi(\varepsilon) = \Pr[\theta|\theta,\varepsilon] - \Pr[\theta|\tau,\varepsilon],\\
&\lim_{\varepsilon\to 0} \phi(\varepsilon) = \sigma,\\
&\phi(0) = \sigma
\end{aligned}$$

defines the *assortment function* $\phi : [0,1) \to [-1,1]$, which is assumed to be continuous, and to converge as $\varepsilon \to 0$ to the matching process's *index of assortativity* $\sigma \in [0,1]$. For uniform random matching, $\sigma = 0$ since $\phi(\varepsilon) = 0$ for all $\varepsilon \in (0,1)$. Alger and Weibull argue that "uniform random matching is unrealistic for most human interactions since it requires that there be zero correlation between the contact pattern that determines how mutations spread in society and the contact pattern that determines who interacts with whom" (2013, p. 2286).

By contrast, there are many natural reasons to expect a positive correlation between such patterns, most notably *homophily*—the tendency of people to interact more with those who are similar to themselves along one or more dimensions (e.g. location, language, culture, profession). There is by now a large literature in sociology and economics on this phenomenon, discussed by Alger and Weibull (2013, §5.2), who also offer a simple model of homophily with unobservable preferences that yields assortativity in the matching process. Moreover, Alger, Weibull, and Lehmann (forthcoming) show that such positive assortativity is generic in a population where individuals stay in the community in which they were born with positive probability. Specifically in the context of money, there is of course a vibrant literature on search as the basis for money's value (see, e.g., Lagos, Rocheteau, and Wright 2017), and in particular criticism of random matching (Howitt 2005; Prescott 2005) over something more assortative.

An individual is a κ-*homo moralis* (Alger and Weibull 2013) if his utility function is of the form

$$u_\kappa(x,y) = (1-\kappa)\cdot\pi(x,y) + \kappa\cdot\pi(x,x)$$

for some $\kappa \in [0,1]$, his degree of morality. If $\kappa = 0$, we have the usual case of "pure selfishness", which Alger and Weibull term *homo oeconomicus*. A type θ belonging to the unit interval will henceforth refer to homo moralis with that degree of morality.

Proposition 1 *If each agent i is a κ_i-homo moralis with $\kappa_i \in (0,1]$, then a BNE of $\mathscr{G}$ implements a maximal competitive equilibrium of $\mathscr{E}$.*

Proof. With both agents being of the same type $\kappa_i \in \Theta$, and hence using the same strategy x,

$$\begin{aligned}
&\arg\max_{z\in X} \Pr[\theta|\theta,\varepsilon]\cdot u_{\kappa_i}(z,x) + \Pr[\tau|\theta,\varepsilon]\cdot u_{\kappa_i}(z,x)\\
&\qquad = \arg\max_{z\in X} u_{\kappa_i}(z,x)\\
&\qquad = \arg\max_{z\in X} \left((1-\kappa_i)\cdot\pi(z,x) + \kappa_i\cdot\pi(z,z)\right)\\
&\qquad = \arg\max_{z\in X} \left((1-\kappa_i)\cdot V(z,\hat{p}(x)) + \kappa_i\cdot V(z,\hat{p}(z))\right), \qquad (7)
\end{aligned}$$

The strategy x solves (7) if and only if $(x(A), x(B), \hat{p}(x))$ is a maximal competitive equilibrium of $\mathscr{E}$, from which the result follows. The second term of (7) is crucial for this, as it requires that the solution set a market-clearing price that maximizes the resultant expected discounted payoffs across the two possible agent roles. ■

Thus, only a small degree of Kantian motivations is required for selection of a maximal competitive equilibrium of $\mathscr{E}$. Whilst the degree of morality κ_i is left free in Proposition 1, Theorem 1 of Alger and Weibull (2013) establishes that setting it equal to the index of assortativity σ yields evolutionary stability, as the following result captures.

Corollary 1 *If $\sigma > 0$ and $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, then σ-homo moralis is evolutionarily stable against all $\tau \notin \Theta_\sigma$, and implements a maximal competitive equilibrium.*

Whilst the evolutionary stability of σ-homo moralis here is unsurprising in light of Alger and Weibull (2013), its implementation of a maximal competitive equilibrium is more so. In particular, this differs sharply from Alger and Weibull's prediction that the efficient equilibrium of ordinary coordination games is unique only if σ is large enough: In such games, an inefficient equilibrium exists for low enough values of σ because the individual then attaches weight to the payoff she would forgo by not coordinating with the opponent's play. By contrast, in the exchange-economy setting of this paper, any given individual is a price-taker, and hence forgoes no payoff by deviating from a nonmonetary equilibrium to a strategy whose universal adoption would enable trade. An agent with homo oeconomicus preferences, for instance, would be indifferent to deviating to such a strategy, and hence an infinitesimal value of σ is sufficient to make a σ-homo moralis strictly prefer the deviating strategy.

What does this tell us about the stable utility functions, beyond their inclusion of σ-homo moralis? The MIU approach to monetary economics of course has the utility function depend directly on money. Under an appropriate choice of this function, such an individual will belong to Θ_σ; hence, if a maximal competitive equilibrium of $\mathscr{E}$ requires agents to demand non-zero holdings of money, there must be stable utility functions that depend positively on those money holdings in the manner of the MIU approach. Whilst a range of utility functions may thus theoretically be stable, a recent paper by van Leeuwen, Alger, and Weibull (2019) begins the task of estimating the social preferences that emerge under preference evolution.

Example In Sargent's (1987, §6.6) example from the previous section, $\beta_\sigma(x)$ is a singleton for σ-homo moralis for the unique $x \in X_\sigma$ and $\sigma > 0$, giving the maximal monetary equilibrium choice. Hence, by Corollary 1, σ-homo moralis is evolutionarily stable against all types $\tau \notin \Theta_\sigma$, and moreover it implements the monetary equilibrium. ■

It is natural to wonder if we can go further and destabilize homo oeconomicus, but this does not seem immediately possible. Alger and Weibull's Corollary 2 provides sufficient conditions for a type's evolutionary instability, but to invoke it we would need homo oeconomicus to have a unique resident strategy; since it can sustain multiple equilibria, this cannot be the case. However, a population of individuals programmed to play just the nonmonetary equilibrium would certainly be unstable by Alger and Weibull's Corollary 2. Thus, whilst the survival of homo oeconomicus cannot be ruled out, it is inconsistent with the play of just the nonmonetary equilibrium. Moreover, the presence of homo moralis is sufficient to give money value, and hence to select the monetary equilibrium.

A related idea found in the existing literature is that fiat money is given its value by a small subset of agents, specifically the government, who accept fiat money and hence coordinate play on a monetary equilibrium (e.g. Obstfeld and Rogoff 1983; Ritter 1995; Aiyagari and Wallace 1997; Li and Wright 1998). But whilst the behavior of these agents is an assumption in such papers, here I have shown it to be a natural evolutionary consequence of assortative matching, and not confined to governmental actors.

References

Aiyagari, S. R., and N. Wallace. 1997. "Government Transaction Policy, the Medium of Exchange, and Welfare". *Journal of Economic Theory* 74:1–18.

Alger, I., and J. W. Weibull. 2013. "Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching". *Econometrica* 81:2269–2302.

— . 2016. "Evolution and Kantian Morality". *Games and Economic Behavior* 98:56–67.

Alger, I., J. W. Weibull, and L. Lehmann. Forthcoming. "Evolution of Preferences in Group-Structured Populations: Genes, Guns, and Culture". *Journal of Economic Theory*.

Araujo, L., and B. Guimaraes. 2014. "Coordination in the Use of Money". *Journal of Monetary Economics* 64:38–46.

Bergstrom, T. 2003. "The Algebra of Assortative Encounters and the Evolution of Cooperation". *International Game Theory Review* 5:211–228.

Bewley, T. F. 1980. "The Optimum Quantity of Money". In *Models of Monetary Economies*, ed. by J. H. Kareken and N. Wallace, 169–210. Minneapolis: Federal Reserve Bank of Minneapolis.

Blume, L., and D. Easley. 1990. "Implementation of Walrasian Expectations Equilibria". *Journal of Economic Theory* 51:207–227.

Cass, D., and K. Shell. 1980. "In Defense of a Basic Approach". In *Models of Monetary Economies*, ed. by J. H. Kareken and N. Wallace, 251–260. Minneapolis: Federal Reserve Bank of Minneapolis.

Debreu, G. 1970. "Economies with a Finite Set of Equilibria". *Econometrica* 38:387–392.

Dekel, E., J. C. Ely, and O. Yilankaya. 2007. "Evolution of Preferences". *Review of Economic Studies* 74:685–704.

Ely, J. C., and O. Yilankaya. 2001. "Nash Equilibrium and the Evolution of Preferences". *Journal of Economic Theory* 97:255–272.

Grandmont, J. M. 1974. "On the Short Run Equilibrium in a Monetary Economy". In *Allocation Under Certainty: Equilibrium and Optimality*, ed. by J. H. Dreze. New York: John Wiley & Sons.

Groves, T., and J. Ledyard. 1987. "Incentive Compatibility Since 1972". In *Information, Incentives and Economic Mechanisms*, ed. by T. Groves, R. Radner, and S. Reiter, 198–204. Minnesota: University of Minnesota Press.

Güth, W., and M. E. Yaari. 1992. "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach". In *Explaining Process and Change*, ed. by U. Witt. Ann Arbor, Michigan: University of Michigan Press.

Hahn, F. H. 1973. "On the Foundations of Monetary Theory". In *Modern Economics*, ed. by M. Parkin and A. Nobay. New York: Barnes & Noble.

Hart, S., and A. Mas-Colell. 2015. "Markets, Correlation, and Regret-Matching". *Games and Economic Behavior* 93:42–58.

Hellwig, M. F. 1993. "The Challenge of Monetary Theory". *European Economic Review* 37:215–242.

Howitt, P. 2005. "Beyond Search: Fiat Money in Organized Exchange". *International Economic Review* 46:405–429.

Hurwicz, L. 1979. "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points". *Review of Economic Studies* 46:217–225.

Hurwicz, L., E. Maskin, and A. Postlewaite. 1995. "Feasible Nash Implementation of Social Choice Rules When the Designer Does Not Know Endowments or Production Sets". In *The Economics of Informational Decentralization: Complexity, Efficiency, Stability*, ed. by J. Ledyard, 367–433. Boston, MA: Springer.

Kiyotaki, N., and R. Wright. 1989. "On Money as a Medium of Exchange". *Journal of Political Economy* 97:927–954.

— . 1993. “A Search-Theoretic Approach to Monetary Economics”. *American Economic Review* 83:63–77.

Kocherlakota, N. R. 1998. “Money is Memory”. *Journal of Economic Theory* 81:232–251.

Lagos, R., G. Rocheteau, and R. Wright. 2017. “Liquidity: A New Monetarist Perspective”. *Journal of Economic Literature* 55:371–440.

Leeper, E. M. 1991. “Equilibria Under ‘Active’ and ‘Passive’ Monetary Policies”. *Journal of Monetary Economics* 27:129–147.

Li, Y., and R. Wright. 1998. “Government Transaction Policy, Media of Exchange, and Prices”. *Journal of Economic Theory* 81:290–313.

Mas-Colell, A. 1985. *The Theory of General Economic Equilibrium: A Differentiable Approach*. Cambridge: Cambridge University Press.

Matsuyama, K. 1991. “Endogenous Price Fluctuations in an Optimizing Model of a Monetary Economy”. *Econometrica* 59:1617–1631.

Obstfeld, M., and K. Rogoff. 1983. “Speculative Hyperinflations in Maximizing Models: Can We Rule Them Out”. *Journal of Political Economy* 91:675–687.

Palfrey, T., and S. Srivastava. 1987. “On Bayesian Implementable Allocations”. *Review of Economic Studies* 54:193–208.

— . 1989. “Implementation with Incomplete Information in Exchange Economies”. *Econometrica* 57:115–134.

Postlewaite, A., and D. Wettstein. 1983. “Continuous and Feasible Implementation”. *Review of Economic Studies* 56:603–611.

Prescott, E. C. 2005. “Comments on ‘Inflation, Output, and Welfare’ by Ricardo Lagos and Guillaume Rocheteau”. *International Economic Review* 46:523–531.

Ritter, J. A. 1995. “The Transition from Barter to Fiat Money”. *American Economic Review* 85:134–149.

Samuelson, P. A. 1958. “An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money”. *Journal of Political Economy* 66:467–482.

Sargent, T. J. 1987. *Dynamic Macroeconomic Theory*. Cambridge, MA: Harvard University Press.

Schmeidler, D. 1980. "Walrasian Analysis Via Strategic Outcome Functions". *Econometrica* 48:1585–1594.

Selten, R. 1980. "A Note on Evolutionarily Stable Strategies in Asymmetric Animal Contests". *Journal of Theoretical Biology* 84:93–101.

Shi, S. 1995. "Money and Prices: A Model of Search and Bargaining". *Journal of Economic Theory* 67:467–496.

Starr, R. M. 1989. "Determinacy of the Price Level, Positivity of the Price of Money". In *General Equilibrium Models of Monetary Economies*, ed. by R. M. Starr, 291–295. San Diego, CA: Academic Press.

Townsend, R. M. 1980. "Models of Money with Spatially Separated Agents". In *Models of Monetary Economies*, ed. by J. H. Kareken and N. Wallace, 265–304. Minneapolis: Federal Reserve Bank of Minneapolis.

Trejos, A., and R. Wright. 1995. "Search, Bargaining, Money, and Prices". *Journal of Political Economy* 103:118–141.

van Leeuwen, B., I. Alger, and J. W. Weibull. 2019. "Estimating Social Preferences and Kantian Morality in Strategic Interactions". Working paper.

Wallace, N. 1980. "The Overlapping Generations Model of Fiat Money". In *Models of Monetary Economics*, ed. by J. H. Kareken and N. Wallace. Minneapolis, MN: Federal Reserve Bank of Minneapolis.

— . 1996. "A Dictum for Monetary Theory". In *Foundations of Research in Economics: How Do Economists Do Economics?*, ed. by S. Medema and W. Samuels. Cheltenham, UK: Edward Elgar Publishing.

Wallace, N., and T. Zhu. 2004. "A Commodity-Money Refinement in Matching Models". *Journal of Economic Theory* 117:246–258.

Wettstein, D. 1990. "Continuous Implementation of Constrained Rational Expectations Equilibria". *Journal of Economic Theory* 52:208–222.

Woodford, M. 1995. "Price Level Determinacy Without Control of a Monetary Aggregate". *Carnegie–Rochester Conference Series on Public Policy* 43:1–46.

Wright, R. 1995. "Search, Evolution, and Money". *Journal of Economic Dynamics and Control* 19:181–206.