A Conceptual Framework for a Theory of Liquidity

James Culham

This thesis is submitted in total fulfillment of the requirements for the degree of Doctor of Philosophy

Federation Business School
Federation University – Australia
PO Box 663
University Drive, Mount Helen
Ballarat, Victoria, 3350
Australia

April 2018

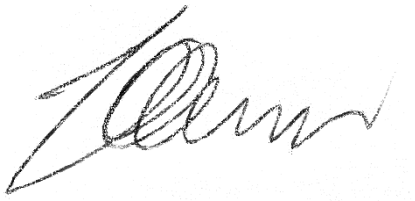
Abstract

This study contributes to the understanding of liquidity in two ways. First, it considers the multifaceted nature of liquidity and its relationship with money. Second, it constructs a conceptual framework for a theory of liquidity. The first contribution is achieved by clarifying and categorising the various forms of liquidity to identify those overlooked by the existing literature. The second contribution consists of a realist critique of the literature on liquidity and money to highlight the strengths and weaknesses of each theoretical approach. The study reflects on the attempts to analyse liquidity using moneyless models of perfect barter with the assumption that every commodity exhibits perfect saleability; an assumption that removes any need for a medium of exchange and, moreover, crowds out all other forms of liquidity. It is concluded that, because liquidity is a social and monetary phenomenon, it cannot be analysed with models populated by a representative agent consuming a single commodity. Furthermore, this conclusion is not altered by the introduction of 'financial frictions', which are fundamentally at odds with the nature of money. Instead, the clarification of the nature of liquidity forms the basis for an interpretation of Keynes's theory of liquidity preference that emphasises its reliance on liquidity in general, not money in particular. The study introduces the terms redemption liquidity and exchange liquidity to explain the trade-off that underpins the theory of liquidity preference. Properly interpreted, the theory of liquidity preference can then address many of the deficiencies prevalent in the dominant theories of the rate of interest. The study therefore has implications for monetary policy and asset pricing.

Statement of Authorship

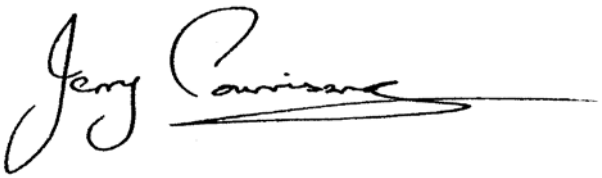
Except where explicit reference is made in the text, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No other person's work has been relied upon or used without due acknowledgement in the text and references of the thesis.

Candidate:



James Culham

Principal Supervisor:



Associate Professor Jerry Courvisanos

Associate Supervisor:

Professor John E. King

Acknowledgements

I owe a great debt of gratitude to Tim Thornton for first suggesting the idea for this project, to John King for leading me to believe that it was possible, and to Jerry Courvisanos, who, as an invaluable guide throughout the whole process, actually made it possible. I must also thank the late Basil Moore for opening my eyes to a branch of monetary literature that most resembled the financial world that I had been working in for over 20 years. Thank you to Colin Rogers for providing the theoretical foundations and much-appreciated encouragement at the beginning of my research journey. I would also like to give my thanks to the various contributors to the economics blogosphere (Perry Mehrling, Nick Rowe and John Cochrane as notables in a list too long to provide in full). They, without always knowing it, challenged my ideas and forced me to think ever more deeply.

Most importantly, I would like to thank my family for their patience and understanding, as my 'aura of unavailability' has meant that I have been missing from some of their important moments. I would also like to recognise the encouragement and interest shown by my friends and colleagues. Finally, I would like to dedicate this work both to my father and to Alan Sayers for the joy that the thought of this project gave them.

Table of Contents

List of Tables and Figures

Part I Introduction

The aim of this research is to contribute to the understanding of the role of liquidity in monetary theory and asset pricing. Problems arise when analysing a monetary economy using the orthodox academic approach: general equilibrium models, being perfect barter, have no fundamental place for money. Since all goods can be readily traded against other goods, money, rather than providing a means for efficient exchange, is entirely unnecessary. Instead, creating a role for money requires introducing it as a friction—the opposite of its observed value to society. Similarly, liquidity—how easily goods can be bought and sold—has no place. To understand the effects of liquidity, and the observed periods of dysfunction when it is missing, requires that money take a realistic place in the theoretical structure. Only then will policy responses to episodes like the Global Financial Crisis 2007-2008 (GFC) be grounded in a realistic assessment of their fundamental cause.

1 Overview

1.1 Introduction

Historically monetary theory has developed along many disconnected lines, with vastly different views on such issues as: the effect of money on the real economy; whether money is best analysed as a medium of exchange or a store of value; whether the money stock is under the control of the central authorities or expands to meet the needs of trade; or whether it is interest rates or the stock of money itself that has the primary monetary effect on economic activity.

Before the GFC, the prevailing orthodox opinion of both practitioners and academics was that the development of the so-called 'Shadow Banking' system[1] was beneficial due to its ability to accelerate risk sharing. In this view, shadow banking, money-market funding, and securitisation[2] were innovations that enabled progress toward the ideal of frictionless markets, in which all resources are allocated efficiently and optimally. Asset pricing, which aims to 'understand the prices or values of claims to uncertain payments' (Cochrane, 2005b, p. xiii), also uses this same frictionless, ideal world of complete markets which allows all contingencies to be insured. Assuming a set of financial securities that covers all future outcomes allows for a tractable solution to the problem of pricing financial options and other exotic derivatives. A fundamental, underlying assumption is that unlimited amounts of all assets can be traded at equilibrium market prices—in other words, perfect liquidity (Nesvetailova, 2010; Mehrling, 2011).

Much has been written in an attempt to identify the causes of the GFC: from a global saving surplus to the testosterone levels in bank traders (Davies, 2010; Rajan, 2010). Liquidity has also been identified as a chief contributor: as traditional banking, with established practices for liquidity management from Bagehot (1873), evolved into the shadow banking system (Mehrling, Pozsar, Sweeney, & Neilson, 2013), the familiar concepts of liquid reserves and

[1] Shadow Banking consists of bank-like activities performed by non-banks, well defined as 'money market funding of capital market lending' (Mehrling et al., 2013, p. 2).

[2] Securitisation is a financial technique for the bundling and selling of otherwise-illiquid assets.

equity-capital buffers were replaced by security repurchase ('repo') agreements[3] and credit default swaps[4], respectively (Mehrling, 2011). Despite some agreement that the 'recent crisis, we all know, was characterized by massive illiquidity' (Tirole, 2011, p. 287), no settled agreement on its cause, or causes, exists (Jefferis, 2017). For example, Gorton and Metrick (2012) consider the GFC to be akin to a classic bank run in the shadow banking system and the repo market in particular, an assessment rejected by Michell (2017) as having both analytical and empirical difficulties. This confusion extends into theory: 'When central bankers discussed liquidity in 2008, they found a neglected territory full of the ruins of old conceptual structures that were not quite inhabitable' (Beggs, 2012, p. 1).

The collapse of the Long-Term Capital Management, a highly leveraged hedge fund, (Lowenstein, 2001) and the GFC, both events where liquidity suddenly vanished, have exposed the effects that a breach of the assumed perfect-liquidity conditions can have on both the financial system and the macro-economy. At the other extreme, post-GFC, many central banks worldwide have conducted enhanced open-market operations (OMO)[5] known as 'quantitative easing' (QE) which go far beyond simply reversing the liquidity shortage which contributed to the crisis. Naturally, this initiative has produced a variety of responses as to its efficacy and predicted effects that reflect the diverse schools of monetary theory (Kregel, 2014; Woodford, 2012).

There is a dissonance: liquidity abundance is needed for financial arbitrage and asset pricing, but liquidity scarcity is needed for price-level determinacy in macroeconomics. Liquidity is often conceptualised as a friction preventing an ideal state of pure exchange, rather than a social benefit created by intermediaries. There is merit in the view that money is not just a special commodity used until the barriers to perfect liquidity are removed; it is a particular form of intermediated wealth claim. Liquidity in general and money in particular are created in response to uncertainty. Although there cannot be an excess of

[3] A repo is a collateralised loan involving a contractual sale and subsequent repurchase of a security, often for as short a period as one day. It provides a low-risk form of very liquid investment for large corporations.

[4] A CDS is a fixed, periodic premium made in exchange for a contingent payment in the event that a named, third-party defaults or restructures its debt.

[5] QE/OMO is the purchase of financial assets, generally government bonds, by the central bank with newly produced central-bank money.

liquidity, since asset prices adjust to the amount that is created, it may well disappear when it is most needed.

The distinction between scarce and abundant liquidity is arbitrary and misleading. Liquidity should not be considered a friction preventing pure exchange, but as a social phenomenon enabling economic agents to cope with uncertainty. As such, its existence is socially beneficial. In brief, liquidity, in various ill-defined and possibly conflicting forms, is recognised as empirically significant in asset pricing, but its theoretical significance is disputed and poorly understood and warrants further study and theoretical analysis.

1.2 The Research Questions

In the context of the problems identified in Section 1.1, the following two research questions have been posed:

1. What is the relationship between liquidity and money in the economics and finance literature and why is this relationship flawed?

2. How can the theory of liquidity be conceptualised in a framework to guide future research on liquidity?

1.3 Significance of the Research

Mainstream monetary theory and general equilibrium asset-pricing theory are both constructed on the assumption of perfect and unlimited liquidity. As Long-Term Capital Management and the GFC have shown, however, a shortage of liquidity can have a potentially disruptive effect on the real economy. Liquidity is subject to the classic fallacy of composition—it is not available when everyone needs it.

This research aims to develop a deeper understanding of liquidity and money and to provide a more realistic framework for evaluating their fundamental causal relationships. This understanding could have policy implications, especially in relation to answering the question of how to construct financial regulations to ensure robustness and prevent financial instability. There needs to be a separate theory of liquidity since ‘it is essential to take liquidity into account in order to discuss any money prices’ (Townshend, 1937, p. 161). The relationship between liquidity and money still remains an open question.

1.4 Structure of the Thesis

The thesis is structured in five parts. Part I continues with Chapter 2 which details the methodology and research method, so chosen to emphasise the realism of the framework. The importance of developing a framework as closely aligned with reality as possible is due to the desire to provide explanations, undercover causality and hence policy implications that are relevant to the world we actually inhabit. Due to the specifications of Chapter 2, an ontological reflection of money, liquidity and banking is conducted in Part II. Part III and Part IV survey the literature associated with Real Analysis and Monetary Analysis, respectively. Existing theories are examined for their realism and historical accuracy; theories based on 'as if' assumptions and purely deductive methodologies chosen for their convenience are rejected. In Part V, a conceptual framework is presented based on a synthesis of the theoretical elements consistent with the research method. The conceptual framework is then used to address the research questions. Chapter 11 concludes by discussing limitations and implications.

1.5 Summary

The conceptual dissonance evident in the literature on liquidity demands that the ontological foundations of the existing theories are uncovered and assessed for their realism. The identification of the causal mechanisms associated with liquidity require a research method grounded in realism, which provides the basis for an immanent critique and revision of existing theories to align better with both the ontology of liquidity and money, and empirical evidence. The result is a clarification of the relationship between liquidity and money and the construction of a conceptual framework for a theory of money. An outline of the framework and its application to questions of policy and implications for future research are given in Chapter 11. The next chapter outlines the methodology and research method selected to ensure the realism of the framework.

2 Methodology and Method

2.1 Introduction

One aim of this study is to address the explanatory failure evident in modern economics (Lawson, 2015, p. 3), especially in relation to the GFC. Since this failure is as much one of methodology as it is of theorising, it is necessary at this point to discuss the methodology and method used in this study and justify their selection.

The many theories of liquidity and money extant in the literature exhibit varying degrees of correspondence with reality and power to explain observed events. It is important, therefore, to adopt research methods that can form a basis for assessing their competing claims. This assessment should be based on each theory's ability to interpret the real world and its usefulness in forming policy statements. Since the aim of the research is to build a conceptual framework for liquidity, the realism of the framework is crucial, as is the identification of the fundamental causal links between liquidity and money. The dissonant views of liquidity and money lead directly to the importance of realism in the research methodology.

The general equilibrium literature, in which money is a friction, does not satisfy this criterion of realism. Instead of an axiomatic deductive methodology with 'as if' assumptions, the realism of the assumptions needs to be verified. Similarly, since money is a social phenomenon, reductionism and methodological individualism are also inappropriate—the system must be considered as being more than simply the sum of its parts. The identification of causal elements requires explanation of the underlying processes, so that instrumentalism, which judges itself via predictive ability alone, is uninformative. The nature of the study also involves unobservable phenomena, such as expected returns, suggesting that positivism, in which only the directly observed is classified as knowledge, is also unsuitable (Boumans & Davis, 2010). Finally, the emphasis on uncertainty, with the associated rejection of Rational Expectations, suggests that a suitable research methodology needs to recognise the complexity and dynamic nature of the underlying economic processes.

2.2 Critical Realism

Critical realism is a research methodology which focuses on causality and the interplay

between reality as it is perceived by the researcher, the analytical framework developed by the researcher, and the descriptive explanations provided by this framework. Reality is assumed to be independent of the researcher, thereby making it possible to uncover causal mechanisms (Elsner, Heinrich, & Schwardt, 2014, p. 543). Thus realism 'asserts that the ultimate objects of scientific investigation exist for the most part quite independent of, or at least prior to, their investigation' (Lawson, 1997, p. 15).

Social phenomena are, by nature, 'intrinsically-meaningful' or 'context-dependent' (Sayer, 1992, p. 29). Although the descriptions of the object under study are not the phenomena themselves, in a social context the nature of object can be affected by the description and meaning given to it. For example, money is concept-dependent:

> Money, and the institutions and practices associated with it, are extremely important in our society ("money makes the world go round!"). A necessary condition of the use of money is that users should have some understanding of what the act of exchanging little metal discs and specially printed pieces of paper for commodities means or "stands for". The users must have some concept of money and also of related phenomena such as rights of ownership, exchange, etc. Hence these social phenomena are "concept-dependent". (Sayer, 1992, p. 30)

It is important to distinguish between the physical behaviour and the meaning of the actions under observation:

> In the case of using money, we could observe the physical behaviour of handing over the little metal discs until the cows came home and we could use every statistical technique in the book to process our observational data, yet if we didn't know the meanings on which the use of money is dependent in the society under study, we would still not have any idea of what was actually happening, or what kind of 'action' it was. (Sayer, 1992, p. 31)

Because of the concept-dependent nature of social objects, it is important to understand, not only the object under study, but also society's understanding of the object: '*in order to understand and explain social phenomena, we cannot avoid evaluating and criticizing societies' own self-understanding*' (Sayer, 1992, p. 39, original emphasis). In this way, false ideas held by agents can be identified as such, even if they form part of the 'common-sense knowledge' (Sayer, 1992, p. 41).

Therefore, the investigation must begin with an 'ontological reflection' (Jespersen, 2009, p. 65), whereby the characteristics of the research subject are identified. Both theorising and the development of knowledge 'require us to "explicate" problematic concepts; that is, give concise definitions to important but vaguely understood terms through re-working their relations with other terms in the network' (Sayer, 1992, p. 81). In the present context, we could postulate that liquidity is one of the 'cases where there are so many competing explications of particularly difficult concepts that it becomes uncertain whether we are still talking about the same thing' (Sayer, 1992, p. 82).

'Liquidity' is a term that has strong metaphorical associations, which have the potential to distort the nature of the observations of liquidity. Theories in the nature of instrumental ordering frameworks ignore 'conceptual problems' or 'problems of meaning', as do deductive mathematical models, which are 'developed and discussed in abstraction from any reference to the real world' (Sayer, 1992, p. 64). Worse still, 'modern economists end up distorting social phenomena just to render them open to treatment by their chosen approach' (Lawson, 2015, p. 122). This distortion of concepts is especially visible in the literature on liquidity, in which it is often analysed by mathematical models with unrealistic assumptions that deduce insights into the behaviour of phenomena that bear little resemblance to more realistic ontological conceptions of liquidity or money.

Much of the literature on liquidity consists of attempts to measure it empirically. These measures and statistical analyses often result in statements of behavioural 'facts', made without allowing for the possibility that the interpretation of data effects the gathering of data. This sharp distinction between facts and theory reflects an inappropriate, theory-neutral view of data, a form of 'naive objectivism' (Sayer, 1992, p. 45), which fails to recognise the interpretation required in the gathering process itself. Realism instead categorises empirical facts as 'thought objects' to avoid confusing them with the 'real objects' to which they refer (Sayer, 1992, p. 47). The perception of the data is concept-dependent, so that observations are 'conceptually-mediated' or 'conceptually-saturated' (Sayer, 1992, p. 54). Care is particularly important in a study of liquidity and money since these concepts are so enmeshed with people's beliefs about them. The phenomena of liquidity and money are, in many respects, simply concepts themselves, and observations of their behaviour will be intertwined with the public's understanding of what they are. We must be careful to think about the hidden concepts, not just with them (Sayer, 1992, p. 52).

Ultimately the research method chosen, if it is to provide policy guidance relevant to the world we actually inhabit, must 'ensure correspondence between reality and theory' (Jespersen, 2009, p. 25). By contrast, Friedman (1953a) recommends an instrumentalist emphasis on predictability over realism of assumptions, which justifies the literature's mathematical models presented as theory along with policy implications, despite explicitly unrealistic assumptions. For Lawson (2015), the problem with this approach lies, not with the unrealistic nature of the assumptions, however, but with the mistaken application of the method of mathematical modelling as a general-purpose tool for investigating social reality. The use of deductive mathematical models relies on the existence and identification of intrinsically stable 'event regularities', or explicit causal links, that are both 'isolatable' and act in such a 'condition of isolation' (Lawson, 2015, p. 15), in which case the system can be described as 'closed'. In an open system, by contrast, the most a researcher can aim to uncover are the 'underlying mechanisms that govern the directly perceivable events and states of affairs of the world' (Lawson, 2015, p. 16). The workings of these mechanisms are observable only as 'tendencies'.

The persistent failure of economic researchers to discover event regularities suggests that the 'nature and conditions of social reality are such that the forms of mathematical deductivist reasoning favoured by modern economists are almost entirely inadequate as tools of insightful social analysis' (Lawson, 2015, p. 109). Models with unrealistic assumptions that merely satisfy the criteria of agreeable conclusions or predictability 'add little to our understanding of social reality' (Lawson, 2015, p. 9) since such a model 'allows for more or less any conclusion to be deduced' (Lawson, 2015, p. 113). Any insights gained are not from the model itself but are 'achieved prior to model construction and incorporated into the modelling process' (Lawson, 2015, p. 8). Instead, social reality needs to be 'understood rather than modelled' (Lawson, 2015, p. 122), which means developing theories that provide more than just an instrumental ordering framework that generates predictions (Sayer, 1992, p. 50). The aim of critical realism is to develop robust, explanatory theories that are 'practically adequate' by satisfying realist criteria for assumptions as well as predictions and insights (Sayer, 1992, p. 70).

The absence of event regularities and difficulties with predicting economic phenomena raise the issue of uncertainty. Since the treatment of uncertainty is as much a methodological question as a theoretical one, some mention of it is made at this point.

Uncertainty is essentially an epistemological problem; it is beyond human knowledge and ability to predict the future with any great precision, and probability theory is a pragmatic attempt to cope with this difficulty. The economics literature makes a distinction between risk and uncertainty, where, in the former, probabilities are known, but not in the latter, so that 'uncertainty corresponds to the situation where knowledge of the probability relation is absent' (Runde, 1990, p. 284). Unlike uncertainty, then, risk is mathematically tractable as 'calculable risk' (Chick, 1983, p. 214) which allows for 'actuarial certainty' (Davidson, 2015, p. 4), and enters the literature as the Rational Expectations Hypothesis, whereby economic agents know the probability distribution of the model itself. Agents with model-consistent expectations then have the ability to construct utility-optimising 'equilibrium' plans that solve the mathematical equations.

The ability of agents to achieve model-consistency requires the stochastic process to have two properties. First, its probability distribution must be stationary, so that its mean and standard deviation are unchanging over time. Second, the process must be discoverable by observation, meaning that a single actual realisation of the process provides sufficient information to estimate the probability distribution with accuracy. Together these properties define an ergodic process.[6] Davidson (1987, 1988, 2015) defines uncertainty as the property of non-ergodicity, because it is 'only in a nonergodic economic world that the concept of uncertainty about the future can be technically defined as differing from risk' (Davidson, 1988, p. 333).

The second ergodic property means that all possible events occur with meaningful frequency, and initial conditions are unimportant or forgotten; there are no 'Black Swans' (Taleb, 2007). An ergodic process is one where time averages suitably represent the key aspects of the probability distribution so that agents can understand the dynamics of an ergodic stochastic process by observing one (sufficiently large) realisation. The estimation that the GFC was an event that 'would happen only once in 5,000 years—that is, it was highly unique' (Davidson, 2015, p. 17) is evidence that the ergodic hypothesis corresponds poorly with the open ontological nature of social reality (Lawson, 2015). The ergodic

[6] In an ergodic process the 'expectation value of the observable is a constant (independent of time), and the finite-time average of the observable converges to this constant with probability one as the averaging time tends to infinity' (Peters & Gell-Mann, 2016, p. 1).

assumption avoids the recognition that economics is a social science that, unlike physics, studies animate objects (Davidson, 1988, p. 332).

O'Donnell (2014, 2016a, 2016b), however, argues that a rejection of the ergodic hypothesis itself requires some form of epistemological certainty, and favours an epistemological definition of uncertainty that 'is independent of whatever deeper stochastic or nonstochastic ontology the world might happen to have' (O'Donnell, 2014, p. 206). Contributing to the difficulty of defining uncertainty is the possibility of chaotic dynamics. Although chaotic dynamics are deterministic, and so predictable, the practical limitations of human knowledge make prediction impossible (Moore, 2006). Furthermore, because chaotic processes are sensitive to initial conditions, the inevitable problems of measurement error make them more and more unpredictable over time. The difficulty with modelling chaotic dynamics is not ontological; it is an epistemological problem of human capability. Because conceptual difficulties arise in agreeing the appropriate treatment of uncertainty as being ontological or epistemological, it is noted here that it is beyond the scope of this study to develop the theory of uncertainty. Only the implications of uncertainty are considered in the context of liquidity, which itself serves as a means of reducing uncertainty.

In conclusion, this study (and hence the choice of research method) is motivated by the contribution that theory can make to explaining social reality. The focus is placed on the ontological nature of the phenomena of liquidity, the alignment of existing theories with this ontology, and the realism of assumptions supporting these theories. By clarifying the concept of liquidity (and, by extension, money) the intention is to increase the explanatory power of economic theory by placing liquidity into a realistic theoretical structure.

2.3 Research Method

At a high level, the application of the critical realist research paradigm involves an iterative process of both induction and deduction, known as retroduction, whereby hypotheses generated by the analytical model are confronted by reality and the insights gained are used to refine and generalise the model. From this process a deeper understanding of the unobservable causal mechanisms in the 'deep stratum' of reality is uncovered (Jespersen, 2009, p. 71).

The emphasis placed on ontology determines much of the research activities. The first stage of the research involves an ontological reflection on the nature of liquidity and money,

using the existing literature as a point of departure. A similar reflection is conducted on the banking system since it is an important institutional element in the provision of liquidity. This ontological reflection is possible by taking the position that reality is independent of the research activity itself but is also informed by 25 years' experience as a practitioner in the financial markets. This direct engagement with the concepts being analysed in their real-world setting imparts a deep understanding from experience that supports the coherence of the ontological reflection.

Once completed, the ontological reflection allows for a critical analysis of the literature on liquidity and money, categorising the essential core of each theory, identifying assumptions and their correspondence with reality. The approach taken is to consider critically each existing theory of liquidity and money and assess its applicability from a realist perspective. Although, on the surface, this approach would seem to be trying to reconcile the incommensurable, it is an essential part of the realist methodology to work with pre-existing theories and look for overlaps:

> If theories are instead thought of as more or less distinctive localities within a continuous conceptual map, which is continually and unevenly evolving, both continuity and novelty and discontinuity can be recognized in the development of knowledge. Some localities may be distant and poorly connected to others, but new links may be established. If we are to avoid the opposite poles of relativism and naïve objectivism, the hermeneutic character of the development of knowledge and the interdependence of sense and reference must be understood. (Sayer, 1992, p. 76)

An attempt is made to discover overlaps in theory, since it is not the case that 'falsification of a part must be fatal to the whole' (Sayer, 1992, p. 73). This quest for synergies should not be overdone but is especially pertinent in the case of interpreting the empirical finance literature, where the theories used in the measurement of the observations are not necessarily the ones needed to interpret the results. Nevertheless, the research effort will focus on the literature where monetary aspects are fundamental to the analysis.

This body of literature and its realist analysis will be used to construct a conceptual framework to understand liquidity and explain its effects on the financial system, asset prices and economic activity. This understanding will lead, in turn, to a theoretical framework for assessing questions of policy and practice.

2.4 Research Design

The structure of the theoretical analysis will proceed along the following route. First, the nature of money and the role of banks in its creation will be established. This ontological platform allows for a taxonomy of liquidity to be offered. The taxonomy frames the subsequent examination and analysis of the existing theories related to liquidity.

The analysis of theory is conducted in two parts. The first part takes the form of an immanent critique of Real Analysis with the focus on general equilibrium theory and its ability to incorporate liquidity and provide insightful analysis with correspondence with reality. Because the effect of liquidity is most significant in the behaviour of asset markets, key aspects of liquidity's relationship with the general equilibrium approach to asset pricing and the associated empirical literature are drawn out.

Then, the next part of the analysis critiques the theories in the tradition of Monetary Analysis, most notably Keynes's theory of liquidity preference, but also other post-Keynesian and heterodox monetary theories. Finally, a synthesis of the analysis of liquidity is presented as a conceptual framework in the penultimate chapter.

2.5 Summary

Events like the GFC provide an empirical justification for the inclusion of liquidity as a fundamental aspect of monetary theory. Also, due to the complexity of the subject, there is a need to establish the ontology of liquidity and money as they exist in reality. In addition, expected returns, a key determinant of asset prices, are unobservable. Critical realism, by providing a framework for uncovering deep causal mechanisms and tendencies, is an appropriate methodology for the research.

So that the research can be built on a realistic platform that allows for policy implications, the ontological notions of liquidity and money must be established. This is the focus of Part II, next.

Part II Ontology

Following the research method described in Chapter 2, the three chapters of Part II conduct an ontological reflection outlining the preliminary characteristics of liquidity, and the related phenomena of money and banking. A map of the ontological landscape of money and banking will provide guidance when assessing the existing definitions and theories of liquidity. Chapter 3 considers the nature of money itself, and the existing theories of money, and a similar exercise is conducted for the banking system in Chapter 4. In Chapter 5, an overview of the liquidity literature yields a taxonomy of liquidity. As stated in Section 2.2, all knowledge is necessarily theory-laden and theoretical statements are unavoidable, so presentations of theory are given when necessary for context. The result is a picture of liquidity that attempts to avoid both *ad hoc* assumptions and appeals to mythical historical processes.

3 The Nature of Money

3.1 Introduction

Money's relationship with liquidity depends on the ontological perception of money. Money is regarded as the most 'liquid' asset, and so it is important to establish, or at least outline the assumptions of, what money 'is' as a precondition for an analysis of liquidity. The purpose of this chapter is to consider whether money has a common essence, or instead whether each form has its own distinctive or institutional-specific features. It is important to clarify the different treatments of money since a 'clear, unambiguous taxonomy is essential for good scientific enquiry' (Davidson, 2006, p. 139). To establish this ontology the approach taken is to assess critically the existing literature, categorising the essential core and identifying the assumptions of each theory to assess their relation with reality. Theories based on 'as if' or unrealistic assumptions, or otherwise lacking explanatory content, will be rejected. In this way, an ontological picture of money is drawn and used to inform the subsequent analysis of liquidity.

Money has taken many forms across history, such as metallic coins, gold standard, fiat currencies and even, supposedly, cows, shells and cigarettes. In looking at the ontology of money, however, it is less important whether money developed spontaneously to solve the problems of barter (Menger 1892; Niehans, 1978; Smith, [1776] 1910), or some other social practice (Graeber 2012; Ingham, 2004a; Martin 2013), what is important is the logical basis of money. We do not necessarily need to consider the origins of money to understand its essence, and the analysis of the 'historical origin of money' does not necessarily reveal 'its nature or logic' (Schumpeter, 1954, p. 289, n. 5). Speculations and evidence concerning the origins of money provide clues as to its nature, and any theory of money must be consistent with the observed characteristics of money even in the early stages of its development. Davies (2002, p. 35) recommends that 'monetary economists should take much more interest in examining primitive money to compensate for their previous neglect' with the result 'that our understanding of modern money would be significantly improved'.

Definitions of money are wide and varied. Some emphasise its legality, stating that money is the 'thing that the state declares will legally discharge any contractual obligation under the civil law of contracts' (Davidson, 2006, p. 140). Others downplay the legal aspect to focus on money's ability to settle tax obligations (Keynes, 1930a, p. 6). The customary

aspect of money can be emphasised, so that money '*is anything that is widely used for making payments and accounting for debts and credits*' (Davies, 2002, p. 29, original emphasis), or 'what is commonly offered or received for the purchase or sale of goods, services or other things' (Galbraith, 1975, p. 15). The third-party nature of money in transactions can be highlighted, whereby money 'may be defined as a commodity or group of commodities customarily paid and received in exchange for other commodities and services without reference to the personal credit of the one who offers it' (Young, [1924] 1999, p. 266).

In addition, there is a distinct, but important, relationship between the unit of account and the money object itself.

> Money itself, namely that by delivery of which debt-contracts and price-contracts are *discharged*, and in the shape of which a store of General Purchasing Power is *held*, derives its character from its relationship to the Money-of-Account, since the debts and prices must first have been expressed in terms of the latter. (Keynes, 1930a, p. 3, original emphasis).

At the outset, it should be noted that this distinction between unit and object, and the implied preconditions of the unit of account for the existence of the object, allows for more than one object or medium of exchange to answer to the description of money (Ingham, 2004a, p. 113).

The separation between the unit and object is not always respected, however. For Einzig (1954, p. 57), 'money may be defined as an object or unit conforming in a reasonable degree to some standard of uniformity, used for reckoning and making payments and accepted with the ultimate intention of using it for making payments', thereby combining and blurring all the concepts. Niehans (1978, pp. 118-119) disregards the logical distinction between unit and object, for if there is no fixed relation between the unit of account and the medium of exchange, and the unit is purely abstract, then it has no economic significance and should be ignored. Instead Niehans prefers the more concrete term 'medium of account' to represent the object which has a price of unity by definition.

As for the money object itself, there are broadly two views. The first, the commodity view, is that the value of money is derived from its metallic or commodity content. The second, the credit view, is that money is essentially tradable credit and is, even in the form of

coinage, merely a token that represents a debt payable. Since commodity money is restricted by the physical stock of the commodity, whereas tokens have no recognisable production function, the commodity view can be taken to express 'the liquidity scarcity view of economics', whereas the credit view adopts 'the liquidity abundance view of finance' (Mehrling, 2000b, p. 15). These views are incompatible and are a source of conceptual dissonance, a problem that must be addressed.

In either view, but with varying degrees of emphasis, three characteristics are commonly assigned to money: unit of account, medium of exchange, and store of value. The first represents money's role as an abstract measure of the size of debt and wealth; the second refers to the use of money as a facilitator for commercial transactions. The third is that agents hold money as a means of obtaining claims to future resources.

In general terms the commodity view is known as 'metallism' and the credit view (especially when it pertains to state money) is known as 'cartalism' (or 'chartalism') (Schumpeter, 1954, p. 288). Metallism and chartalism can be further divided into their theoretical and practical forms. According to theoretical metallism, it is

> …logically essential for money to consist of, or to be "covered" by, some commodity so that the logical source of the exchange value or purchasing power of money is the exchange value or purchasing power of that commodity, considered independently of its monetary role. (Schumpeter, 1954, p. 288)

For metallism, the value of money is solely derived from the commodity that backs it, such as when coins of fixed weight in gold circulate as money. In Schumpeter's view, however, it is self-evident that 'theoretical metallism is untenable' since it is an analytical error to confuse the logical nature of money with its perceived historical origins (1954, p. 289, n. 5). Practical metallism makes the weaker assertion that the 'monetary unit "should" be kept firmly linked to, and freely interchangeable with, a given quantity of some commodity' (Schumpeter, 1954, p. 288). For example, under a gold standard, the value of the monetary notes in circulation is pegged to gold, and gold reserves are kept in central bank vaults to support this value.

Conversely, 'Theoretical and Practical Cartalism may best be defined by the corresponding negatives' (Schumpeter, 1954, p. 288). In other words, for theoretical chartalism, it is not 'logically essential' for money to have metallic content or be easily convertible into a

commodity, nor for practical chartalism ‘should’ it have such content or convertibility (Schumpeter, 1954, p. 288). The prevalence of fiat, or unbacked, currencies in the modern monetary system confirms that it is not practically necessary for money to derive its value from any fixed commodity. As such, fiat money is often characterised as irredeemable or inconvertible (Buiter, 2005).

3.2 Money and Credit

It is a firmly established principle in the monetary literature that there is a distinction between money-proper, which is issued by the state, and credit, which is merely a claim on money (Keynes, 1930a, p. 5; Schumpeter, 1954, p. 717). Credit, in this view, is an innovation that economises on the use of money-proper and thereby stretches the money supply by increasing its velocity of circulation. This distinction is explicit in the framework of Gurley and Shaw (1960) with ‘outside’ money and ‘inside’ money, where the former is ‘a claim held by consumers and firms against government’ and the latter ‘is based on internal debt’ (p. 73). Within this framework, Gurley and Shaw identify and reject what they call the ‘net-money doctrine’, in which only outside money is economically significant. They argue that the consolidation of private securities with their corresponding inside money claims ignores any effect arising from their translation from illiquid primary securities to liquid secondary securities. What remains in the net-money doctrine is the ‘outside’ sector ‘to avoid reverting completely to economic analysis in terms of a barter society where there is neither money nor demand for money, neither bonds nor demand for bonds’ (Gurley & Shaw, 1960, p. 136).

The question of how much importance to place on credit money has implications for monetary theory and especially the analysis of liquidity. In what follows we consider whether the distinction between pure outside money and inside credit money obscures many fundamental aspects of liquidity and money, and, instead, whether insights can be gained by emphasising their commonality. The approach necessitates expanding the definition of money and de-emphasising the importance of state money, since

> …logically, it is by no means clear that the most useful method is to start from the coin—even if, making a concession to realism, we add inconvertible government paper—in order to proceed to the credit transactions of reality. It may be more useful to start from these [credit transactions] in the first place, to look upon

> capitalist finance as a clearing system that cancels claims and debts and carries forward the differences—so that "money" payments come in only as a special case without any particularly fundamental importance. In other words: practically and analytically, a credit theory of money is possibly preferable to a monetary theory of credit. (Schumpeter, 1954, p. 717)

Therefore, a credit theory of money, in which all money is viewed as credit, is preferred to a monetary theory of credit, where money and credit are distinct, and credit merely serves as a promise to pay money. It is important 'to get into the habit of thinking of money as the highest form of credit, if only as a corrective to intellectual habits that come from thinking of credit as an inferior form of money' (Mehrling, 2000c, p. 2, n. 2).

Identifying money as credit, however, requires further clarification since it bears some similarity to the practical metallist view, where money should be a promise to pay a fixed amount of precious metal (Schumpeter, 1954, pp. 289-290, n. 5). The nuance is that the recognition of money as a token is not, in itself, enough to establish a theoretical basis for money, although it can be taken as a basic ontological observation. The starting point for the following discussion is that, although the institutions and forms of money may have changed over time, the nature of money is itself constant. Furthermore, too sharp a distinction between credit and money is misleading. All money is credit regardless of its issuer, metallic content or institutional arrangements. Consideration of the commodity view serves to clarify and strengthen the claim that money is a credit token.

3.3 Commodity View of Money

Money, as it is embedded in the classical commodity-view literature, is a natural invention that reduces the inefficiency of the primitive state of pure barter. Money is simply a natural evolution of the most liquid or saleable commodity reinforced by agent self-interest (Mises, 1953, p. 32). The effect is that 'this difference in saleableness ceases to be altogether gradual, and must be regarded in a certain aspect as something absolute' so that there are 'goods which have become money and goods which have not' (Menger, 1892, p. 252). The unit of account, because it is the measure associated with the liquid commodity, is

inseparable from the commodity itself.[7] A 'medium of exchange derives its usefulness from some sort of imperfection or "friction" in the market' (Niehans, 1978, p. 1), so the 'use of money increases the efficiency of the economy' (Niehans, 1978, p. 2).

Also common in the classical literature is the axiom of monetary neutrality. In 'Of Money', Hume (1742, para. II.III.1) begins with the assertion that 'Money is not, properly speaking, one of the subjects of commerce; but only the instrument which men have agreed upon to facilitate the exchange of one commodity for another.' This assertion leads Hume to deduce that 'it is of no manner of consequence, with regard to the domestic happiness of a state, whether money be in a greater or less quantity' (Hume, 1742, para. II.III.9). By regarding the real economy as separate from the monetary economy, Hume concludes 'that the want of money can never injure any state within itself: For men and commodities are the real strength of any community' (Hume, 1742, para. II.III.20). Money is a neutral 'veil', or more accurately a 'lubricant' (Niehans, 1978, p. 8), whereby 'an exogenous change in its quantity, once all adjustments have run their course, produces a proportional change in all prices, leaving real phenomena unchanged' (Niehans, 1978, p. 2). The classical literature emphasises money's property as a medium of exchange and de-emphasises its role as a store of value. A seller's only intention is to gain money to be a buyer immediately: 'money can serve no other purpose besides purchasing goods' (Smith, [1776] 1910, pp. 384-385). If money is neutral and can be safely ignored in economic analysis, then it may well follow that liquidity can be ignored as well.

It is customary to consider money and coins as conceptually inseparable, with the metallic content of the coins providing their value. As such, the only substantive difference between the metallic money and coins is that coins are stamped to guarantee their metallic content (Hicks, 1989, p. 46). Analyses with the assumption that the value of money is established by its commodity content do not often explicitly distinguish between coinage and its related commodity (for example, see Kiyotaki & Wright, 1989). The outcome is that one commodity becomes unique in that agents will accept it even though they have no immediate need for it, with coins simply a more convenient way of ensuring this

[7] Marx's theory of money most closely aligns to the theoretical metallist perspective and the commodity view of money. Because of Marx's embedded-labour theory of value, money must be a commodity otherwise it cannot be produced and therefore has no value (Nelson, 2005, p. 66).

acceptance. In a commodity view of money, the motivation for the selection of a monetary commodity is to mitigate the inconvenience of barter, thereby directly contradicting available evidence:

> On one thing the experts on primitive money all agree, and this vital agreement transcends their minor differences. Their common belief backed up by the overwhelming tangible evidence of actual types of primitive moneys from all over the world and from the archaeological, literary and linguistic evidence of the ancient world, is that barter was not the main factor in the origins and earliest developments of money. (Davies, 2002, p. 23)

If not in a coin format, however, commodity money is sited in a 'theory of exchange...in which every commodity could serve as a perfect money' (Niehans, 1978, p. 3, n. 5), in other words commodity 'money' is no more than an untransformed commodity.

Without dwelling too much on the question of the origins of money, we can use some elements of the historic record to eliminate some unnecessary conjectures as to the nature of money. First, the variation in coin size and metallic content was too great for the value of the coin to be based on metal alone (Davies, 2002, p. 75; Ingham, 2004a, pp. 109-110; Innes, 1913, p. 380). Second, many coins were minted without any numerical denomination so that their value in the unit of account could be changed arbitrarily by the issuer as a form of tax (Davies, 2002, pp. 98-99; Innes, 1913, p. 385). Third, the quantity of coins in circulation was too small for common use (Davies, 2002, p. 77; Innes, 1913, p. 389); and fourth, in many cases the value of coins was often too great for day-to-day use (Tymoigne & Wray, 2006, p. 7). Fifth, there were many types of coin, many issued by merchants so that often 'eighty different coinages' were in circulation (Innes, 1913, p. 385; see also Ingham, 2004a, p. 111). Together these imply the existence of other, more common, forms of payment such as credit. An undue focus on state money and the state versus credit-money dichotomy is unwarranted.

Consider the relationship between coins and their metallic content. By the force of arbitrage coins cannot circulate below their intrinsic value; the metallic content of coins provides a lower bound to their value. The fact that they generally circulate above the value of their embedded metal is evidence that they represent more than their metallic content; in other words, they are 'tokens'. The value of coins, even with consistent weights of precious metals, is not due to their metallic content. When metal is minted into coins the difference

between the value of the metallic raw material and the coin created from it, known as *seigniorage*, is profit available to the minting authority. This profit is only available if the value of coins is greater than their value in metal, and provides the motive for private minters (Tymoigne & Wray, 2006, p. 10). 'The Mint price was established with reference to a money of account, and it could diverge, and very often did, from the market price of the metal' (De Cecco, 1991, p. 315). Alternatively, coins that were worth less in their monetary form than in their metallic form would not remain in circulation since they could be melted down and sold as metal for a greater value. Similarly, attempts to maintain a bimetallist standard often failed when the relative market values of the two commodities deviated enough from the official rates of parity to offer arbitrage opportunities (De Cecco, 1991, p. 322; Galbraith, 1975, pp. 78-79).

Historically, and hence theoretically, there is no fixed relationship between the unit of account, coinage and the metallic content of the coinage (Ingham, 2004a, p. 110).[8] The implications for monetary theory are that coins must be deemphasised lest they dominate the conceptual landscape. We must recognise that 'coinage never played any considerable part in commerce, that the monetary unit was distinct from the coinage and that the price of gold and silver fluctuated constantly in terms of that unit' (Innes, 1913, p. 390). The use of credit denominated in a unit of account pre-dates coins by at least two thousand years (Ingham, 2004a, p. 46; Tymoigne & Wray, 2006, p. 7). Historically, clearing of credits has been conducted without coins (Wray, 2004, p. 11), so it is necessary for monetary theory to account for non-money forms of settlement.

Theoretically, coins are a pure expression of practical metallism, and the various forms of metallic standard were important as a cure either for mismanagement or for limiting the *seigniorage* of the state. Mismanagement generally meant that the issued money would not be acceptable in payment of taxes or that the state would succumb to temptations to 'cry down' the currency, with the result that 'the gold standard may have been desirable in an era of monarchs who mismanaged the monetary system' (Tymoigne & Wray, 2006, p.11). Reforms enforcing a specific metallic content of coins were a means of reducing the

[8] The logical separation of the unit of account and coinage is a rejection of a key element of Marx's theory of money: that gold is the evolved form of money and the measure of value must be associated with its social cost of production (Foley, 2005, p. 42).

sovereign's taxation ability so that 'he would be compelled to resort openly to his fiscal powers' (De Cecco, 1991, p. 316). The continuous tug-of-war between the state and its subjects over the metallic content of the state's coinage, and the occasional victory by those in favour of a pure metallic standard, should not obscure the fundamental credit nature of coinage. A closer look at the workings of the gold standard further reveals the fundamental credit nature of money.

3.4 Gold Standard

For a more evident example of a practical metallism, consider the gold standard. Under a gold standard, central banks provide convertibility of their notes into gold at fixed gold points or prices, thereby setting the monetary value of their paper currency equivalent to that of a specific weight of metal. For instance, from 1919 until early 1933 the value of the US dollar was defined so that one ounce of gold was equivalent to $20.67. Thereafter it was raised incrementally by arbitrary decree to finish at approximately $35.00 per ounce early in 1934 (Galbraith, 1975, pp. 223-224). To maintain this fixed-price conversion the central bank maintains a stock of gold in reserve, but the amount of currency issued is not limited to the amount of gold reserves, as in a pure commodity currency. Instead, the amount of currency outstanding is based on the demand for currency at the gold-exchange price and the central bank's open-market operations (Glasner, 1985, p. 47), with the supply of money responding to the demand for money (Glasner, 2000, p. 45, n. 3).

Thus, under a gold standard, a central bank's note is a promise to pay gold. If the flow of redemptions of money for gold were to become too large, thereby diminishing its reserves, the central bank would need to reverse this flow, and the instrument for maintaining the target stock of gold is the central bank's discount rate—raising it reduces the outflow of gold (Keynes, 1930b, p. 303). The value of a currency on a gold standard is pegged by arbitrage to the value of the gold commodity, as an imposed, but not intrinsic, value. The backing of the notes is maintained by use of interest rates as the means to discourage redemption into gold. As a form of open-market operation, raising the discount rate reduces the stock of notes on issue by making it less rewarding for the public to swap non-gold assets with the central bank for notes. The reduction in notes restores the target note-to-gold ratio.

Furthermore, under a gold standard world prices and interest rates are set by international

arbitrage assuming the 'unity of world markets'. Gold flows simply satisfy the demand for money, by a decrease either in the gold-to-money ratio or in the propensity for imports. A decrease in import propensity leads to increased gold imports, which are then converted into money (Glasner, 1985; McCloskey & Zecher, [1976] 1997). Changes in the stock of gold at each central bank do not alter world prices, since, with fixed exchange rates, the vector of world prices is fixed by international arbitrage and the law of one price[9] (Glasner, 1985, p. 56; see also Samuelson, [1980] 2015; and Cottrell, 1997, for a rebuttal). World stocks of gold determine the spot value of the currency, but open-market operations are essential in maintaining and managing its convertibility. In periods when the gold standard was in operation the difficulties in maintaining sufficient reserves to support the promise of convertibility affected the stability of the entire international monetary system, especially in the decentralised US banking system before the introduction of the Federal Reserve (De Cecco, 1991; Mehrling, 2011).

The issuing of paper money in return for gold is conceptually the same as coin minting, except that the monetary token is not based physically on the gold itself. The gold standard clarifies the nature of the gold specie: the central bank holds the physical gold, and issues paper tokens, thereby eliminating the inefficiency of transporting gold for domestic transactions. In the UK in the early 1960s the Royal Mint, a government department, issued coins and the Bank of England issued notes convertible into these coins. The Bank of England would use its notes to purchase coins directly from the Royal Mint, which would use these notes to purchase the metal and labour to create the coins. The two forms of money, state money and central bank money, were distinct, the latter viewed as a promise to pay the former (Sayers, 1964, p. 82). This distinction between state money and central bank money is more likely to be overlooked in the current institutional framework where central bank money is more dominant, and no longer convertible into any other state money.

Although there are historical periods when money forms have been linked to metal, physically or in value terms, the basic nature of money is that it is effectively a credit token whose value is not derived solely from any metallic content. Unless converted into monetary form, silver and gold are commodities like any other, and have no intrinsic

[9] Not to be mistaken for purchasing power parity, which operates at the overall price level and not at the level of individual commodities as does the law of one price (Glasner, 2000, p. 47).

monetary nature.

3.5 Quantity Theory of Money

The ontology of money affects the treatment of money within the monetary literature. Transcending all theories is the quantity theory of money, which has dominated the monetary literature, in one form or another, since the classical period of economic analysis. Because of its dominance, some comment must be made about the quantity theory of money at this point in the study.

Since the time of classical political economy, it has been postulated that the effect of money is 'that the prices of every thing depend on the proportion between commodities and money' (Hume, 1742, para. II.III.12). The result is the quantity theory of money (QTM), based on the equation of exchange:

$$PT = MV. \tag{3.1}$$

By inferring a right-to-left direction of causality, it is claimed that the price level P is determined by the transactions demand T for a fixed stock of commodity money M circulating at a particular 'velocity' V. The QTM embeds the dominant metaphor of money as a veil which obscures, but does not affect, the workings of the real economy (Sowell, 1974, p. 42), and the 'classical dichotomy' (Patinkin, 1965) between the real and the monetary economy. An increase in the money supply could lower interest rates and stimulate economic activity in the short term, but it would have no long-term effects apart from on the price level (Sowell, 1974, pp. 56-57). A surplus of money raises prices because it first raises demand (Patinkin, 1965).

The QTM relies on the independence of the elements comprising the equation of exchange (Schumpeter, 1954, p. 703), so that, most notably, the supply of money is unaffected by the demand for money. Instead, any discrepancies are resolved by adjustments in income in the short term, and prices in the long term. Hence, the theoretical power of the equation of exchange relies on the stability of velocity and its independence from the money supply, but both are undermined 'because of the ease of response of the money supply to changes in the volume of money transactions makes changes in velocity superfluous' (Kaldor, 1982, p. 29). Townshend (1937, p. 161) observes that any amount of money can support any price level depending on its velocity of circulation regardless of its institutional limitations.

Nevertheless, the QTM is maintained by its seemingly unassailable empirical support. 'The central predictions of the quantity theory are that, in the long run, money growth should be neutral in its effects on the growth rate of production and should affect the inflation rate on a one-for-one basis' (Lucas, 1996, p. 665). The strong correlation between price changes and money supply changes are viewed as validating this prediction; the QTM has remarkable success empirically (Lucas, 2014). The QTM 'applies, with remarkable success, to co-movements in money and prices generated in complicated, real-world circumstances' (Lucas, 1996, p. 666), even though it is possible to identify historical episodes that contradict the hypothesis (Schwartz, 1973, p. 249).

Whatever its empirical basis, however, there are logical problems. The 'quantity theory' should be known as 'the quantity theorem', since 'it is not a complete theory of money but merely a proposition about the exchange value of money' (Schumpeter, 1954, p. 312). If we assume a pure metallist standard then any increase in the monetary commodity has its effect via the shape of its supply and demand schedules, just like any other commodity. There is no reason to appeal to the QTM at all. In fact, the theories are conflicting since the QTM implies a proportionate change in price to the change in the monetary commodity's supply, a result that is difficult to achieve for a commodity with an elastic production function. Although the QTM is often associated with the commodity theory of money, it can only apply to inconvertible fiat money (Glasner, 1985, p. 56; Niehans, 1978, pp. 9, 113, 147-149, 193). Fiat money is considered unbacked and is assumed to be accepted in payment simply from the authority of the state, either by legal tender laws or some indefinite form of trust. As such, the QTM is more in alignment with a chartalist view of money:

> This logical affinity of the quantity theorem with theoretical cartalism should be borne in mind: the theorem essentially amounts to treating money not as a commodity but as a voucher for buying goods, though not everyone who does consider money in this light need accept the specific schema offered by the quantity theorem. (Schumpeter, 1954, p. 313)

Fiat money, or base money as it is known in the modern context, is defined as 'an object that is intrinsically worthless (does not appear in any utility or production function) and inconvertible (is not a redeemable claim to something that does, such as a stock certificate)' (Kiyotaki & Wright, 1989, p. 941, n. 11).

> When a commodity is accepted in trade not to be consumed or used in production, but to be used to facilitate further trade, it becomes a medium of exchange and is called commodity money. If an object with no intrinsic value becomes a medium of exchange, it is called fiat money. (Kiyotaki & Wright, 1989, p. 929)

The supposedly irredeemable nature of fiat money is essential to recent proposals to increase aggregate demand by means of 'helicopter' money, with the recommendation that the central bank dramatically increase the stock of fiat base money. Since fiat money is viewed as irredeemable and is always considered an asset by its owner, this helicopter drop is an increase in 'net wealth for the private sector' (Buiter, 2014, p. 3), as per the net-money doctrine. In effect, fiat or outside money is treated as if it were metallic or commodity money, but this treatment is logically inconsistent. Instead, fiat money should be considered another form of bank money with 'negligible costs of production' (Rogers, 1989, p. 172).

Furthermore, the association between fiat money and commodity money is based on a misunderstanding of commodity money; that is, once metal has taken a monetary token form, it ceases to be valued as a metal and takes on the value of a credit token. Fiat money is issued by central banks, which, despite their elevated position in the financial system, are banks both logically and historically. Inconvertible fiat money is a liability, in many respects, like a stock certificate. Stock certificates are inconvertible: they only provide the holder to a claim on the profits of the business but no actual rights to the assets. A limited liability company is an infinitely lived entity until insolvency or shareholders vote to liquidate it. The inconvertible aspect of the definition of fiat money is not strong enough to distinguish it from other liabilities.

Mehrling provides a useful summary of the confusion generated by the introduction of fiat money and the attempt to maintain the commodity view:

> Even today, the notion that currency is in some sense backed by the public credit remains controversial. Most people seem to be more comfortable thinking about currency as "fiat", a kind of paper token taking the place of outside gold, than as a liability of the central bank. (Mehrling, 2000c, p. 7, n. 5)

Mehrling (2000c, p. 7, n. 5) goes on to reject an appeal to legal tender and state authority as a suitable substitute by implying a 'commitment to pay one's bills quite apart from one's actual ability to do so'. The problem is that 'the loss of a commodity anchor seems to have

left us intellectually unanchored as well' (Mehrling, 2000c, p. 7, n. 5).

Instead, the gold standard, correctly interpreted, provides a basis for understanding fiat money, since

> …national money today is best understood as a promise to pay, just as it was under the gold standard. The payor is much the same. What has changed is that what is to be paid is not one particular commodity but rather abstract value itself, the value of the unit of account named on the bill. (Mehrling, 2000c, p. 7, n. 5)

Although the link between the money issued by the central bank and any explicit commodity has been severed, support for the currency is still based on credit-worthiness and the ability to maintain its value through some form of asset backing or offsetting credit flows. Instead, the unit of account has become more important, with the value of money the result of social and political conflict (Galbraith, 1975; pp. 316-317; Ingham, 2004a, pp. 80-85).

Finally, the assumption of money neutrality fails to find support from the commodity view. Commodities have a cost of production and hence 'the neutrality proposition, strictly speaking, does not apply to commodity money (like gold coins)' or 'paper money convertible into gold', with the result that 'the neutrality proposition [only] relates to inconvertible paper currency' (Niehans, 1978, p. 9). The assumption of money neutrality logically and necessarily restricts any associated analysis to that of fiat money.

In summary, the QTM epitomises the unresolved tension between the treatment of money as a commodity and its recognisable modern fiat form. Each aspect is emphasised in different contexts but otherwise remains as a form of conceptual dissonance. Whatever truth is expressed in the QTM, it cannot be satisfactory as an explanation of money, it can only provide an 'as if' theory sustained as one of the 'theories which for the time being have not been bettered' (Sayer, 1992, p. 71). The quantity theory of money stands or falls on its restricted applicability to fiat money.

3.6 Credit View of Money

At the other extreme from the commodity view of metallism is the credit view of money, which asserts that all money is credit (or debt) since any 'sale and purchase is the exchange of a commodity for a credit' (Innes, 1914, p. 152). Note that, in the context of the credit

view of money, the terms ‘credit’ and ‘debt’ can be used interchangeably. Credit money is a debt on the liability side of the issuer’s balance sheet which must correspond, in existence but not necessarily value, to the credit on the asset side of the money owner’s balance sheet. An essential implication is that a money asset cannot exist without a corresponding debt liability on another agent’s balance sheet. Under a credit view of money, fiat money without a corresponding liability does not conceptually exist.

In considering commercial transactions, Hicks (1989, p. 42) insists that they consist of three parts: ‘first is the contract between the parties, consisting of a promise to deliver and a promise to pay (both are needed to make even a constituent part of a transaction); the second and third consist of actual delivery, one way and the other.’ The credit view departs from Hicks by treating the exchange as final even without the resolution of the promise to pay:

> [W]hen a seller has delivered the commodity bought and has accepted an acknowledgment of debt from the purchaser, the transaction is complete, the payment of the purchase is final; and the new relationship which arises between the seller and the purchaser, the creditor and debtor, is distinct from the sale and purchase. (Innes, 1913, p. 394)

For exchange to take place, an acknowledgement of the associated or corresponding debt and credit is enough. This acknowledgment can take a physical form, such as bills of exchange (Sayers, 1964, pp. 46-48), tally sticks (Davies, 2002, pp. 147-153), or abstract financial securities (Lozano, 2015). Settlement, where the credit or promise to pay is later honoured with another commodity, is a separate exchange independent from original. This independence allows the acknowledgement of debt to be transferred to another agent entirely unrelated to the original exchange.

The credit view postulates that money is founded on tradable credit and agrees with the theoretical chartalist position that the value of money is not necessarily derived from any metallic content, and that money is by nature a ‘token’. Importantly, the unit of account and the monetary token are independent. The existence of an abstract unit of account allows for the development of double-entry bookkeeping, which in turn allows for complex debit and credit relations, the interplay of balance sheets and financial engineering. It is money’s use for settlement that makes the difference between debt and money, and explains the departure from and then return to Hicks’s conception of a transaction: credit that can be used in settlement of other credits is money.

The assertion that all money is credit is not to claim, as Goodhart (2005a, p. 760, original emphasis) does, that this 'implies that *all* credit is money'. Likewise, the claim by Bell (2001, p. 150) that 'when a buyer (debtor) and a seller (creditor) enter into a forward contract, money (or equivalently, credit and debt) is created' cannot be substantiated. In many cases, debt is just debt, and debt must be settled by money (Keynes, 1930a, p. 3). Transferability for settlement is what transforms debt into money. Not all credit is equal; only some credit can become money, and, even then, with different degrees of acceptability:

> All money is debt in so far as issuers promise to accept their own money for *any* debt payment by *any* bearer of the money. The credibility of the promises forms a hierarchy of moneys that have degrees of acceptability. The state's sovereign issue of liabilities usually occupies the top place, as these are accepted in payment of taxes. (Ingham, 2004a, p. 198, original emphasis)

Mehrling (2013a) expands on this idea of a hierarchy of money, as a form of pyramid, with (in the example given) gold at the apex, then descending through state money, bank money, down to more modern forms of liquidity such as security repurchase agreements (repo), a key feature of the shadow banking system (Gabor & Vestergaard, 2016). The hierarchy is dynamic in terms of order and quantity of each money, but: 'Always and everywhere, monetary systems are hierarchical' (Mehrling, 2013a, p. 1). It is because of the familiarity of the current hierarchy, or perhaps unfamiliarity with historical versions, that it is customary to distinguish sharply between state money, which is deemed to be money-proper, and credit money. Government money has become so prevalent, that we are in the habit of associating it alone with 'money' (Innes, 1914, p. 152). Because of its direct relation with taxes, state money is currently at the top of the pyramid of money (Bell, 2001, p. 160), but historical evidence shows that other debt tokens have a right to the label of 'money'. Although it is important to recognise the essential hierarchy of money, it is also important that '*the different means of payments may, on a certain level of abstraction, be treated as essentially alike*' (Schumpeter, 1954, p. 719, original emphasis).

As Minsky ([1986] 2008, p. 255) notes, 'everyone can create money; the problem is to get it accepted'[10], and that 'usually there is a hierarchy of monies, with special money

[10] Bell (2001, p. 150) uses this statement to imply that the mere acceptance of another's debts is sufficient for

instruments for different purposes'. Ultimately, the hierarchy is a form of Gresham's Law, where bad money drives out good. Higher forms of money in the hierarchy are hoarded where possible and, lower forms of money in the hierarchy are preferred for transactions, and hence substitutes come into existence (Galbraith, 1975, p. 56).

As mentioned previously (see Section 3.3), recent studies compiling anthropological research into the origins of money provide no empirical support for the commodity view of money assumed in the classical and neoclassical literature.[11] Instead, the evidence shows that credit is the more predominant form of payment throughout history, with metallic coins playing a much smaller, reduced role (Graeber, 2012; Ingham, 2004a; Martin, 2013). 'There is no question but that credit is far older than cash' (Innes, 1913, p. 396).

> Innes's version of the credit theory of money has, then, three main elements. First, money is primarily an abstract measurement of value. Second, all forms of money are credit in that their value consists in their ability to redeem a debt; "money" cannot exist without the existence of a debt to be redeemed. Third, credit instruments predate coined currency and historically represent the major form that money has taken. (Ingham, 2004b, pp. 178-179)

According to the credit view, metallic coins, which have led to the development of the commodity view of money, are merely tokens used for the purpose of paying tax, with a 'necessarily close connection between the minting of money and collecting it back in taxes. Minting and taxing were two sides of the same coin of royal prerogative, or, we would say, monetary and fiscal policies were inextricably interconnected' (Davies, 2002, p. 147). The exchange value of coins would always exceed their intrinsic value as base metal, otherwise no metal would be taken to the minters and the existing stock of coins would be melted

---

the creation of 'money', seemingly because, since money is a balance-sheet relation, all balance-sheet relations are money. The importance of transferability for settlement is not recognised by Bell as the distinction between debt and money. This omission distorts her subsequent discussion on the hierarchy of money, which is more in terms of market liquidity than money due to the confusion between money and debt. This confusion results in state bonds, which cannot be used as a medium of exchange, incorrectly appearing in the hierarchy of money (Bell, 2001, p. 159, n. 2).

[11] This alleviates Keynes's concerns, in reviewing Innes (1913), that the credit theory is 'much diminished by an entire absence of any references to authorities' (Keynes, 1914, p. 420).

down. The value of the coins would depend, not on their commodity content, but on the creditworthiness of the issuer, namely the state and its taxation ability. Thus

> …the value of credit or money does not depend on the value of any metal or metals, but on the right which the creditor acquires to "payment," that is to say, to satisfaction for the credit, and on the obligation of the debtor to "pay" his debt, and conversely on the right of the debtor to release himself from his debt by the tender of an equivalent debt owed by the creditor, and the obligation of the creditor to accept this tender in satisfaction of his credit. (Innes, 1914, p. 152)

Whereas the conventional view is that 'credit is the lending out of the money commodity' (Martin, 2013, p. 14), the credit view of money shows that credit is delayed settlement, where the settlement can be achieved, not just by money or commodities, but by other, third-party credits. The existence of third-party credits acceptable as payment is the social invention of money. In the process of commerce, agents accumulate credits, which only after a period are discharged by some form of commodity or settlement. 'To focus on the commodity payment rather than the system of credit and clearing behind it was to get things completely the wrong way round [sic]' (Martin, 2013, p. 12). Credit, and its role as a medium of exchange, plays a fundamental role in commerce throughout history.

A credit-money economy often involves three parties in a transaction: buyer, seller and the money issuer (Rossi, 2012, p. 81; Rochon & Rossi, 2013). In this case, paying with a third party's promise to pay settles the transaction for the buyer, so that by the transferability of money, an agent can settle outstanding debts with the credits of other agents (Ingham, 2004a, p. 72). A failure to maintain the balance of current credits and debts affects the current value of the obligor's credit:

> The debts which count in the depreciation of the monetary unit are those which are contracted without any provision for their payment and which are either payable at sight as in the case of currency notes or payable at short terms and have to be constantly renewed for want of credits with which to cancel them. (Innes, 1914, p. 157)

Thus, the 'law of the equation of debts and credits' states that 'the value of a credit on any debtor depends on an equation between the amount of debt immediately payable by the debtor credit and the amount of credits which he has immediately available for the

cancellation of his debts' (Innes, 1914, p. 160). Although third-party credit money may circulate through the hands of many agents to enable settlement, its value derives from the fact that it will eventually find its way back to the original issuer as payment of a debt. All agents must eventually make some form of repayment, and this repayment generally takes the form of handing over the credits of a third party. In the modern economy, that third party is the banking system, and modern commerce is conducted by the transferral of bank deposits.

The recognition that money is credit emphasises the intermediated and social nature of money; it is not just an asset to the holder but, also a claim on a borrower. In the commodity view, money is similarly an asset to the holder, but has no liability counterpart, and hence is unambiguously wealth. If money is a credit and not a commodity, then all money in existence must have an associated asset and liability entry on the money issuer's balance sheet. Not only are government bonds not net wealth (Barro, 1974), but neither is government money. Money, at a macroeconomic level at least, is indicative of past investment, and it follows that the hoarding of money cannot be a leakage from investment. The idea that money hoards are 'spending power in some absolute sense removed from the economy' (Turner, 2013, p. 5) can only be aligned with the commodity view of money. Instead, money hoards are the result of an asset allocation decision because the investment decision associated with its existence has already been made. That money must indicate investment, and not hoarding, is an important detail that we return to when we consider the 'loanable funds' theory of the rate of interest in Section 6.2.

Next, we must consider the nature of the unit of account, which as Keynes (1930a, p. 3) suggests, is a pre-condition for money, before clarifying the relationship between money and the state.

3.7 Unit of Account

Without a unit of account distinct from monetary forms, 'we have scarcely emerged from the stage of barter' (Keynes, 1930a, p. 3). The existence of a monetary medium of exchange implies the existence of a unit of account, but not the converse (Hicks, 1989, p. 43). It must be emphasised again that the term 'money' represents at least two quite distinct concepts. The first is the unit of account, which is an abstract measurement of debt and value, no different in essence from a measure of weight or length. The second includes the objects

that perform the function of discharging a debt or acting as a store of value. The use of the term ‘money’ blurs this distinction: for instance, Hicks (1989, p. 43) cites the historical example of cattle used as ‘money’, but the evidence suggests that cattle were a unit of account, not a means of payment (Davies, 2002, p. 42). As we will see with the term ‘liquidity’, the term ‘money’ fails to provide the ‘clear, unambiguous taxonomy’ (Davidson, 2006, p. 139) required for analysis. In fact, in common usage ‘money’ often refers to wealth more generally.

Sayers (1960, p. 711) finds the idea of the abstract unit of account less ambiguous than the object of money. The money object itself, which has the functions of medium of exchange and store of value, is much more ambiguous: ‘I suspect that, because money in the abstract sense is an unambiguous concept, people have subconsciously believed that there must also be a simple answer to the question of what is money in the concrete sense.’ To maintain clarity, it must be recognised that the term ‘money’ does not refer to a singular thing with both the characteristics of a unit of account and a medium of exchange; instead, these are entirely different concepts. The unit of account is important: it measures all transactions and wealth. The money object itself has a supporting role, transformed from any metallic content it may have.

As a pre-condition to the existence of money objects and markets, the unit of account must be introduced by an authority (Ingham, 2004a, p. 49). Tymoigne and Wray (2006, p. 3) ‘believe that the monetary unit almost certainly required and requires some sort of authority to give it force’ and, rather than arising spontaneously, ‘that the unit of account was socially determined rather than the result of individual optimization.’ The importance of the unit of account, imposed by some authority, must be emphasised. ‘Without this unit of account, no debt instruments could have become monetary instruments because they could not have been recorded in a generalized unit of account but rather only as a specific debt’ (Tymoigne & Wray, 2006, p.12). In all likelihood, the unit of account is derived from measures of value associated with fines, tributes, ransom, or bribery (Davies, 2002, pp. 25-26; Einzig, 1954, p. 62; Graeber, 2012; Ingham, 2004a, pp. 90-93). The ‘money of account is logically anterior and historically prior to market exchange and market value’ (Ingham, 2004a, p. 179).

Whether or not the abstractness of the unit of account is important analytically (Niehans, 1978, pp. 118-119), we know that historically there are examples of this phenomenon.

Instead of the necessity of an object that corresponds to the unit of money, Davies (2002, p. 29, original emphasis) recounts that 'for around half the long monetary history of the £ sterling in Britain this was not the case: there was no such *thing* as a pound; it existed only as a unit of account. There are numerous similar examples.' Just so, other instances of currencies as a unit of account without an associated physical object are the euro currency (Ingham, 2004a, p. 6) and its predecessor the European currency unit, or ECU, (Stigum, 1990, p. 212), and the URV ('unit of real value') during the introduction of the Brazilian *Real* (Sachs & Zini, 1996). The fact that modern money often only exists as computer entries is entirely consistent with the logical separation of unit and object: 'the dollar is a measure of the value of all commodities, but is itself not a commodity, nor can it be embodied in any commodity. It is intangible, immaterial, abstract' (Innes, 1914, p. 159).

It has been recognised that, logically, money is credit denominated in a unit of account, imposed by an authority or convention. The power to set the unit of measure, however, does not award control over all that can be measured. The limitations of the authority and its theoretical relationship with money must be considered next.

3.8 State Money

Knapp's (1924) state theory of money places emphasis on the government's power to declare what it will accept in payment of taxes and the related concept of legal tender. In opposition to the commodity view of Menger (1892) that money emerged spontaneously from barter, the state theory denies the existence of money without state involvement, such that 'Money is the Creature of Law' (Schumpeter, 1954, p. 1090). The theory is based on the premise that, since money cannot arise spontaneously as in the commodity-money view, it must have arisen from government intervention. The result is a 'non-market based theory of money' which 'places the state centre-stage', wherein the ability to tax is necessary and sufficient to give value to government currency (Bell, 2000, p. 154). Consequently, the state theory is presented as more general than the 'most traded' commodity theory (Bell, 2000, p. 160, n. 1).

The state theory of money has been revived by neo-chartalists or proponents of 'Modern Monetary Theory'. At its core, the state theory claims that the government is the 'monopoly supplier' of money and, since it has the power to issue inconvertible fiat or outside money, it does not face a hard budget, or finance, constraint (Bell, 2000, 2001; Tymoigne & Wray,

2013; Wray, 1998, pp. 74-96).[12]

> The monetarily sovereign government is the monopoly supplier of its currency and can issue currency of any denomination in physical or non-physical forms. As such the government has an unlimited capacity to pay for the things it wishes to purchase and to fulfill promised future payments, and has an unlimited ability to provide funds to the other sectors. Thus, insolvency and bankruptcy of this government is not possible. It can always pay. (Tymoigne & Wray, 2013, p. 5)

In this view, government is the sole issuer of currency and therefore government spending, which creates this currency, must logically occur before the public is able to pay taxes. The government's budget constraint, in which government spending must be financed by a combination of taxes, borrowing and new money stock, is rejected (Tymoigne & Wray, 2013, p. 6). Instead, it is claimed, the government can achieve all of its expenditure with new currency issues; only thereafter does it need to adjust the level of taxation to ensure that that stock of money is consistent with price stability.

Based on the essential elements of the nature of money and the credit view outlined up until this point, several problems with the state theory can be recognised. The first observation is that the theory does not distinguish between the unit of account, which indeed must be imposed by custom or authority, and the medium of exchange, which can take many forms. It is not logically correct to observe that the state has the ability to dictate the unit of account and then conclude that state money is the only form of money, any more than to conclude that the state is the only agent that can create objects that can be measured in metres or kilograms. This assertion will only hold if state money is the only form of money that can be used in transactions, a claim that has no historical or logical accuracy. As Ingham (2004a, p. 179) points out, 'economic value is not natural like the relatively constant properties of, say, distance and weight.'

The historical support of the assertion that state money is the sole form of money, is undermined by state money's limited use in comparison with credit, which has been the

[12] Neo-Chartalist theory, restricted to the special case where the state pursues a flexible exchange-rate policy, is packaged into 'Modern Monetary Theory' (MMT) with other ideas such as 'Employer of Last Resort' and 'Net Financial Assets', which are outside the scope of consideration here; see Wray (1998), Tymoigne & Wray (2013). For critiques, see Lavoie (2013), Mehrling (2000a) and Palley (2015).

predominate form of payment. ‘So unimportant indeed was the coinage that sometimes Kings did not hesitate to call it all in for re-minting and re-issue and still commerce went on just the same’ (Innes, 1913, p. 389). In addition, government money was often only accepted at a discount to other forms of money, it did not even maintain its face value in the nominated unit of account (Davis, 1900, p. 16; Innes, 1914, p. 153). ‘For most of economic history, the state has been rather weak and consequently dependent on money markets and financial markets for its projects’ (Mehrling, 2000c, p. 7). A monetary theory that can account for historical episodes must deal with the observed episodes of weakness of state money (Beggs, 2016). ‘The main problem for the state, in most periods and most places, has been to find a way to insert itself into the ongoing hierarchy of private credit and international money’ (Mehrling, 2000c, pp. 7-8). The same problems of establishing the worth of state credit are visible in attempts to adopt currency pegs or boards (Mehrling, 2000c, p. 13).

Ultimately, being a monopoly supplier of its own money does not logically make the state the monopoly supplier of all money. In the presence of other forms of money, such as private or merchant money, the balance between its spending and taxation will determine the strength of demand for state money. It does not follow from the ability to produce unlimited credits, that the state (or anyone) has unlimited spending capacity. Under the credit view, it is also undoubtedly true that each agent is a monopoly supplier of its own money, and the state is no exception. The state can emit as much credit, in either money or debt form, as will be accepted by the other agents in the economy. Nevertheless, there are limits, since the

> …view of the role of the state as the sole creator and guarantor of money, although useful as a corrective to the metallistic theories current at the end of the nineteenth century, nevertheless carries the state theory of money to an absurd extreme. (Davies, 2002, p. 26)

One problem with the state theory stems from its failure to recognise other forms of money; even the difference between state and central bank money is overlooked. When other monetary forms are recognised, they are viewed through lens of the state theory. The claim that only taxes give value to money is even applied to bank money. Rather than realising the bank money is necessary as part of the banking system’s role as a clearing house of credit, Bell (2001, p. 159) asserts that ‘even if convertibility to state money were suspended

indefinitely (except at clearing houses), bank promises, as long as they were accepted in payment of taxes, would continue to accepted.' Although taxes are sufficient to give state money value (Tymoigne & Wray, 2013, p. 10), it does appear that, in the state theory, only taxes can give money value and, since only the state can raise taxes, only state money has value. This claim must be rejected historically and logically.

Next, it can be recognised that the state has the ability to impose tax obligations on all other agents, with the result that these agents are obliged to obtain state credits to discharge their tax obligations. The necessity to pay tax with a state-issued token can give value to state money[13]; without the ability to tax, the value would fall to zero (Bell, 2001, p. 155, n. 2). The actual value, however, is not determined, nor is it the only way in which a money token can evoke value. In knocking over a metallist straw man, neo-chartalism overreaches by claiming not just 'that state determines the money of the economy by declaring what it will accept in payment to itself' (Bell, 2001, p. 154), but also that the state need not 'pay market prices for the goods and services it purchases' (Bell, 2001, p. 156, n. 2).

It is because of its inability to determine the actual value of money that Schumpeter also finds Knapp's (1924) state theory of money 'absurd' (Schumpeter, 1954, p. 1091) since it provides no theory or explanation of value:

> Had Knapp merely asserted that the state may declare an object or warrant or ticket or token (bearing a sign) to be lawful money and that a proclamation to this effect or even a proclamation to the effect that a certain pay-token or ticket will be accepted in discharge of taxes must go a long way toward imparting some value to that pay-token or ticket, he would have asserted a truth but a platitudinous one. Had he asserted that such action of the state will *determine* the value of that pay-token or ticket, he would have asserted an interesting but false proposition. (Schumpeter, 1954, p. 1090, original emphasis)

The ability to impose taxes requiring repayment using state credit is sufficient to give state credits value, but it is not sufficient to give the state arbitrary power to determine the value of its credits.

---

[13] Cochrane (2005c) outlines a fiscal theory of the price level explaining the value of money by the ability of the government to tax. See Buiter (2005) and Rogers (2006) for a critique.

Similarly, Goodhart (2005a), whose 'judgement is that the chartalist/credit approach is historically valid' (p. 760), nevertheless doubts the neo-chartalists' claims that 'the state can choose anything it likes to function as the "money thing"' (Wray, 2004, p. 243), beyond setting the unit of account. The neo-chartalists appear to reject the requirement that the state, in issuing credits backed by taxes, must maintain the value of its credit, thus directly contradicting a core tenet of the credit theory of money:

> The value of a credit depends not on the existence of any gold or silver or other property behind it, but solely on the "solvency" of the debtor, and that depends solely on whether, when the debt becomes due, he in his turn has sufficient credits on others to set off against his debts. If the debtor neither possesses nor can acquire credits which can be offset against his debts, then the possession of those debts is of no value to the creditors who own them. (Innes, 1913, p. 393)

Nevertheless, the claim that taxes are not necessary to 'finance' government spending (Bell, 2000) is consistent with the principles of the credit theory. Like all agents, governments can spend via emitting credit and this credit may or may not then take the form of money or debt. It is also consistent with credit theory that taxes are the source of the offsetting credits required to maintain the value of government credit 'as an asset on the government's balance sheet' (Mehrling, 2000a, p. 401), and taxes do not need to be received prior to government spending. Nevertheless, the government, like any agent, has a budget constraint, whereby its expenditures are met by emitting new debts, and old debts must have means of retirement if they are to maintain their value. Tymoigne and Wray (2013, p. 6) claim that the budget constraint is 'not an equation describing the choices to fund government expenditures.' This claim is incorrect.

Other attempts by the neo-chartalists to enlist the support of the credit theory in their quest to accentuate the usefulness of a state-issued fiat currency (Wray, 2004) are not entirely successful. Innes himself explicitly points to the fallacy of their approach: 'of all the false ideas current on the subject of money none is more harmful than that which attributes to the government the special function of monopolising the issues of money' (Innes, 1914, p. 152). Innes is very clear that all money is credit that credit takes many forms, and arises in any transaction. State money is a 'promise to pay' just like any other and, in credit theory terms, the emphasis placed on state money by the neo-chartalists is unwarranted and unnecessarily myopic. 'All forms of money are identical in their nature' (Innes, 1914, p.

154), so instead of focusing on the irredeemable nature of modern central bank money, the credit view emphasises the similarities between state money and other forms of credit. The former is a 'promise to pay' based on the government's ability to impose a taxation debt on the public, whereas the latter are valued by their use in repaying private debts.

In the credit view of money, state or outside money receives no special treatment; it is credit like all others, including those historical periods where the monetary system was based on a gold or silver standard. The ability to tax does not excuse the state from problems of maintaining the value of its debt:

> The fact, however, is that the more government money there is in circulation, the poorer we are. Of all the principles which we may learn from the credit theory, none is more important than this, and until we have thoroughly digested it we are not in a position to enact sound currency laws. (Innes, 1914, p. 161)

The credit theory rejects the net-money doctrine, where outside money is unquestionably considered to be net wealth to the public. Although it is often overlooked in the literature, neo-chartalism 'has a natural affinity with' the quantity theory of money (Mehrling, 2000a, p. 405) and is therefore essentially a form of monetarism (Davies, 2002, p. 26), even though many advocates of MMT would be uncomfortable with this affinity. Since it is restricted to state money only (Bell, 2000, p. 154), it is also more limited than the credit theory. In the neo-chartalist view, state money is the only money available, and the only consequence of its over-issue can be inflation. By assuming away other forms of money, the neo-chartalist theory is left by default, at the expense of 'conflating money and state finance' (Mehrling, 2000a, p. 401).

Instead, state or outside money, although having the appearance of never 'becoming due', may lose its acceptability in payment if its value is not maintained. 'The fact that the modern state arrogates to itself the right to determine what is and what is not money does not give it the alchemical power to create something from nothing' (Mehrling, 2000a, p. 401). Instead, fiat money is simply another form of inside money: 'the liability of the central bank' (Mehrling, 2000a, p. 401). Conceptually, '"fiat money" is very much like metallic money subjected to a 100 per cent seigniorage charge' (Young, [1924] 1999, p. 271); and *seigniorage* is ultimately linked to the ability to impose taxes.

3.9 Summary

All money is credit, including state money, and its value is ultimately derived from the debt it represents. Even fiat money issued by a central bank, which appears to be irredeemable, is credit since it ultimately is used to repay tax obligations or is an intermediated claim on the central bank's assets. In addition, a distinction must be made between the money issued by a central bank and money issued by the state, and indeed the other forms of money.

The state theory of money, being a special case of the credit theory, is true in its limited realm but should otherwise be subsumed into the broader and more applicable credit theory. The recognition that money is credit allows the emphasis to move from state money to credit more generally, since state money is but one form of credit. The current institutional arrangements mean that state money is generally the strongest form of money but this strength is not a continuously-observed phenomenon throughout history, and so should not be allowed to dominate the analysis of money. State money should not take such a '*logically* privileged position' (Schumpeter, 1954, p. 719, original emphasis). In the analysis of liquidity, all forms of money need attention so that the picture of liquidity is complete.

All money is credit in that it ultimately represents a promise to pay and is in essence an intermediated claim on the credit of the money issuer. Even where the monetary asset is explicitly linked to a fixed amount of commodity, the ultimate product is a promise to pay the stated amount of commodity. Monetary theories in which a commodity itself is the medium of exchange are non-monetary; instead, they represent a model of idealised pure commodity exchange. Although there are many examples of metallic-based currencies, the value of money does not arise from any metallic content through any logical necessity. Consequently, monetary theories where an unrefined commodity is the basis of money can be rejected on grounds that their ontology is insufficiently aligned with reality.

Money is recognised as the most-liquid asset, and so a broader definition of money allows for a richer investigation of liquidity. As will be seen in Chapter 5, 'liquidity' is also a term that covers a host of concepts without sharp boundaries, and by considering all forms of 'money' more of the concept of liquidity can be encompassed. By treating all money as credit, we are able to increase the boundaries of the object of study so that all forms of liquidity can be captured in the analytical framework. For instance, recognising the distinct nature of the unit of account allows for the analysis of the shadow banking system as a

provider of modern money. Complementing this approach is a rejection of the quantity theory of money because of its limited application to fiat money only.

In summary, the evidence suggests that money and credit are more complex phenomena than is admitted by the commodity or state theories. The realisation that money is a special form of credit in a dynamic hierarchy invites us to picture it as a spectrum rather than in a fixed supply, and to reconsider money's neutrality in the acts of commerce.

For a deeper understanding of the credit nature of the monetary system the primary supplier of money in the modern economy—the banking system—is considered in the next chapter.

4 Banks – Clearing Houses of Credit

4.1 Introduction

The characteristics of the banking system, which plays such a large part in the creation and management of money, are fundamental to the understanding of liquidity and money. This chapter places the phenomenon of money into its natural institutional context: the banking system.

A bank is fundamentally a 'dealer in debts or credits' (Hawtrey, 1919, p. 4) or equivalently 'an institution that deals in debts' (Young, [1924] 1999, p. 272). The activity of banking is best considered as a swapping of IOUs, where the IOUs issued by the banking system are themselves used as a medium of exchange (Mehrling, 2011, p. 72; Tymoigne & Wray, 2006, p. 13). The constant issuing and cancelling of debts are represented by the terms efflux and reflux, and signify that money, in the form of bank liabilities, is continuously created and destroyed. By the reflux mechanism any excess in the supply of money is automatically extinguished (Lavoie, 1999). 'Debts and credits are perpetually trying to get into touch with one another, so that they may be written off against each other, and it is the business of the banker to bring them together' (Innes, 1913, p. 402).

Since bank liabilities are money, it follows that banks do not simply re-lend state money as suggested by the misleading analogy that money is deposited in a bank just like 'we entrust our bag to the cloakroom clerk' (Cannan, 1921, p. 29). Instead,

> …[the banker] lends the same sums over and over again *before the first borrower has repaid*: that is to say, he does not merely find successive employments for the sum entrusted to him, but many employments which that sum then fills simultaneously. (Schumpeter, 1954, p. 320, original emphasis)

The cloakroom analogy fails because the ticket received in exchange for a bag cannot be used as a bag. The ticket received in exchange for state money deposited at a bank, on the other hand, can be spent as if it were itself state money. The purchasing power embedded in the original money can be 'used' twice, something not possible with bags. Instead, it must be recognised that a bank can create money by the expansion of its balance sheet: 'from the point of view of the rest of the economy, the bank has "created" money. This is not to be denied' (Hicks, 1989, p. 58).

> Banking is not money lending; to lend, a money lender must have money. The fundamental banking activity is accepting, that is, guaranteeing that some party is creditworthy. A bank, by accepting a debt instrument, agrees to make specified payments if the debtor will not or cannot. Such an accepted or endorsed note can then be sold in the open market. A bank loan is equivalent to a bank's buying a note that it has accepted. (Minsky, [1986] 2008, p. 256)

Banks do not 'lend' money; it is the distinguishing feature of banks that their liabilities are money:

> It should be obvious that banks' role in money creation does not arise from their use of "leverage" (debt financing) per se. All sorts of firms, commercial and financial alike, employ leverage—often in large amounts—but are not said to create money. Nor is money creation synonymous with "lending" (or investing) in and of itself. Finance companies are prominent in the lending business, after all. Rather, it is the combination of banks' investment activities and their reliance on a particular *type* of liability—deposits—that is responsible for their role in augmenting the money supply. (Ricks, 2011, p. 77, original emphasis)

Through the lens of the commodity view, however, this money creation process is often regarded as 'alchemy' (King, 2016). The idea that bank credit economises money-proper by increasing its velocity is an impediment to a deeper understanding of credit and money. 'There surely must be other ways of expressing these practices than by calling these bank notes embodiments of velocity of circulation—a velocity so great that it enables a thing to be in different places at the same time' (Schumpeter, 1954, p. 320). Rather than increasing the velocity of money, banks increase the supply of the means of payment—a feat that cannot be achieved with any commodity. Instead, money should be identified with credit since

> …credit instruments, or some of them, intrude into the monetary system; and, by the same token, money in turn is but a credit instrument, a claim to the only final means of payment, the consumers' good. By now this theory [the credit theory of money]—which of course is capable of taking many forms and stands in need of many elaborations—may be said to prevail. (Schumpeter, 1954, p. 321)

Consequently, the class of money should be extended to include banknotes and checking

deposits, and ‘be broadly defined to include all debts that transfer purchasing power from the future to the present’ (Wray, 1992b, p. 301). Next, we consider one of the most important credit instruments.

4.2 The Bill of Exchange

Essential to understanding banking and credit clearing is the bill of exchange or trade credit. In this primitive instrument, all modern instruments affecting the provision of liquidity can be recognised and the decoupling of the state from the supply of liquidity can be seen in its simplest form. The bill of exchange allows trade to be enacted without the need for any recognisable form of money. In addition, a bill of exchange can be transferred and used as a secondary medium of exchange, so that bills form both part of the supply and the demand for money as not just ‘elements in the total demand for money but elements in its supply’ (Schumpeter, 1954, p. 695). The bill of exchange stands as a form of private money (Ingham, 2004a, p. 116).

Furthermore, the bill of exchange has a long historical pedigree:

> The use of bank deposits, bills of exchange and other forms of credit as substitutes for banknotes had long been acknowledged in the literature; on the historical order of events as we know them, it might rather have been argued that banknotes were a substitute for bills of exchange. (Sayers, 1960, p. 717)

According to Davies (2002, p. 34), banking in Babylon predated coinage by close to a millennium. Analysis of the bill of exchange, or trade credit, is useful to complete the picture of money as payment in exchange.

A bill of exchange can most easily be visualised as a post-dated cheque, whereby final payment is not due until the maturity of the bill, that is, the post-date. At the beginning of the process, the creditor ‘draws’ the bill on the debtor, the drawee. Then the drawer sends the bill to the drawee, who ‘accepts’ it, thereby committing to paying the owner of the bill the notional amount at its maturity. The drawer need not hold the bill to maturity. Instead, the bill can be sold, generally for a sum less than its notional face value, by means of ‘discounting’, effectively paying the purchaser an interest rate. Even after selling the bill, the original creditor is also liable for the notional amount of the bill should the original drawee fail to pay (Sayers, 1964, pp. 46-48). There is no limit to the amount of rediscounting possible for a single bill. Bills of exchange that are hoarded and held to

maturity can be classified as ‘simple credit’ to distinguish them from a ‘currency of credit’ that circulates and acts as payment in other, unrelated transactions (Mehrling, 1996, p. 334). Since bank liabilities are monetary and private non-banking liabilities are not, the discounting of bills by banks provides liquidity and maturity transformation services, an activity that is not naturally provided by other economic agents.

The process of discounting bills of exchange is the logical and historical starting point for understanding the business of banking. Bills of exchange, as a form of credit to make payment in transactions, can arise simply in the normal course of trade. Even a ‘market for *acceptances of bills of exchange*’ (Hicks, 1989, p. 48, original emphasis) does not necessarily require specialists in financial dealing such as bankers and brokers. Nevertheless, from such a market for bills, it is natural that discount specialists would emerge (Hicks, 1989, p. 49). These specialists are dealers that supply ‘money’ in the present for a price, the interest rate. Bills of the same quality but different maturities have different interest rates due to ‘a pure matter of time-preference—a pure rate of interest’ (Hicks, 1989, p. 49). Similarly, different borrowers, even with the same maturity and nominal amount in the unit of account, will be worth different amounts because of credit risk and lack of familiarity across all other agents.

Intermediaries are essential for a competitive market in loans due to the problem of familiarity with borrowers, otherwise ‘the establishment of a competitive market for simple lending is not at all a simple matter’ (Hicks, 1989, p. 50). The banking system, by collecting the bills of many agents, can balance debts and credits, and provide a system of cancelling or ‘clearing’ offsetting bills. ‘There is thus a constant circulation of debts and credits through the medium of the banker who brings them together and clears them as the debts fall due’ (Innes, 1913, p. 403).

> Demand deposits have exchange value because a multitude of debtors to banks have outstanding debts that call for the payment of demand deposits to banks. These debtors will work and sell goods or financial instruments to get demand deposits. The exchange value of deposits is determined by the demands of debtors for deposits needed to fulfill their commitments. (Minsky, [1986] 2008, p. 258)

Ultimately, credits are useful only because they provide a means of repaying a debt to the original borrower. State money repays a debt to the state. Despite appearances, bank debts are not repaid with state money; a debtor to a bank can only repay with an equivalent

amount of bank money. From the point of view of a bank, state money does not cancel a bank debt; it merely replaces it with a credit against the state.

In considering the relationship between the state and the banking system, Tymoigne and Wray (2006) observe that the state spends its liabilities through the banking system, conceptually just like any other economic agent. They also claim that 'the state cannot "spend" its tax receipts, which are just reductions of outstanding state liabilities' (Tymoigne & Wray, 2006, p.13). There is no reason, however, to exclude the possibility that the state can amass credits against private agents, denominated in the unit of account, held as liabilities of the banking system, just like any other economic agent. The state has the power to get its liabilities accepted as credit in the banking system because of its ability to impose taxes, and this is generally, not always, the better credit in the hierarchy (Mehrling, 2013a), but it does not follow that it is somehow any different in concept from other money or credit issuers. State money itself can be viewed as a form of bill of exchange. The most significant difference from a standard bill of exchange is that state money is undated and requires no discount on acceptance. It can be used in the clearing process to settle a tax obligation immediately. As we shall see, this undated nature of money is crucial.

Hicks (1989, p. 56) assigns three activities to a bank: '(1) accepting deposits, (2) discounting bills, and (3) making advances to customers.' Once it is recognised that central bank or state money is simply a state-issued bill of exchange, then activities 1 and 2 differ only in the identification of the issuer of the bill being discounted. Accepting deposits is conceptually the same as discounting bills of exchange issued by the state; the outcome is that the bank intermediates lending to the state or central bank, which was formerly lent directly by the depositor. In either case, the lending has already been made, since the money exists, and the existence of a new deposit does not imply that the bank can lend this again. Nor does a bank require a reserve of state money to engage in more lending. Accepting deposits should be interpreted as intermediating government borrowing, nothing more.

Similarly, Hicks's third activity—making advances to customers—is merely the simultaneous act of creating a bill of exchange and discounting it. The act of taking a loan from a bank is the simultaneous creation of a new bill drawn by the bank on the borrower and a new bill drawn by the borrower on the bank, the latter being bank money. The borrower uses the bill drawn on the bank for the purchase requiring the loan. In essence, Hicks's three activities are only of one genus: the creation and clearing of credit.

Furthermore, potential borrowers can also pre-arrange lines of credit, or overdrafts, whereby this process of dual bill creation can occur at the discretion of the borrower (Moore, 1988). The existence of undrawn overdrafts, and their implications for the measured stock of money, is revealed by a comparison between the traditional, post-war English system where overdrafts were left undrawn and the contemporary American system where they were fully drawn and the unused balances were left on deposit (Sayers, 1964, p. 35). Conceptually these two forms of unused spending power are the same, but the implications for the money stock are significant and highlight the difficulties associated with its measurement in the quantity theory (Black 1970b; Schumpeter, 1954, p. 705). In the former setting the money stock, as measured by deposits, is smaller, yet there is no difference in the amount of 'money', or purchasing power, available. 'Banks create credit and thus purchasing power' (Turner, 2013, p. 20) but 'any attempt to measure the definitive quantity of transaction balances, separate from other bank (or shadow bank) liabilities has become increasingly impossible' (Turner, 2013, p. 22).

The implications for bank liquidity are equally significant. In fact, the very existence of the bill of exchange in its undiscounted form is a monetary element that needs consideration. Trade credit 'introduces considerable elasticity into the response of business to efforts by the monetary authorities to compress liquidity' (Radcliffe Committee, 1959, p. 103). The use of trade credit is an extension of money achieved by non-financial firms expanding their balance sheets, and these liabilities should be included in the class of liquid assets. 'The distinction between banks as creators of credit and other firms as users or intermediaries in the monetary field is if not completely false at least misleading' (Sayers, 1960, p. 713).

> As soon as we realize that there is no essential difference between those forms of "paper credit" that are used for paying and lending, and that demand, supported by "credit," acts upon prices in essentially the same manner as does demand supported by legal tender, we are on the way toward a serviceable theory of the credit structure and, in particular, toward the discovery of the relations between prices and interest. (Schumpeter, 1954, pp. 718-719)

For the purposes of effective demand, the relevant quantity is the availability of the means of purchasing power, in any form. Once circulating and non-circulating credit is included in the means of payments for transactions, it can be shown that the means can expand *pari*

passu with the value of transactions (Mehrling, 1996, p. 333).

> It is this wide concept of liquid assets that we must put, in the place conventionally occupied by "the supply of money," as the monetary quantity influencing total effective demand for goods and services. And we must interpret it widely enough to include credit that can be brought into existence concurrently with a decision to exercise demand. (Sayers, 1960, p. 712)

Trade credit in a closed cycle of efflux and reflux still provides effective demand, even though the net amount may be zero (Sayers, 1960, p. 713).[14] Its effect is significant since 'trade credit is so large in relation to bank credit that a comparatively small lengthening of trade credit would normally offset quite a large proportionate reduction in bank credit' (Radcliffe Committee, 1959, p. 103). The banking system is not the only source of credit creation, control over credit supplied by banks cannot be justified unless a firm relationship exists between bank and non-bank credit (Sayers, 1960, p. 714). Is there any limit to this credit creation?

4.3 The Creation of Credit

The discounting or acquisition of bills of exchange by banks is 'effected chiefly by the creation of liabilities in the form of deposits' (Dunbar, 1922, p. 33); no particular reserve of state money is required. This naturally leads to the question:

> What determines the limit to which this process can be carried?
>
> If depositors seldom demanded the payment to which they are entitled, and were content with the mere transfer of their existing rights among themselves as a conventional currency, the bank might dispense with holding any large amount of specie or cash in any form and keep most of its resources employed in its productive securities. (Dunbar, 1922, p. 33)

In response, it must first be recognised that the possibility of an unlimited expansion of the banking system as a whole, as distinct from the constraints on individual banks, only exists

[14] The reader should be aware that Professor R. S. Sayers (Sayers, 1960, 1964) was also a member of the 'Committee on the Working of the Monetary System' chaired by Lord Radcliffe, and a principal author of the committee's report (Radcliffe Committee, 1959).

if the banks offer indistinguishable monies (Glasner, 1985, p. 57). If each bank promises to convert their monetary liabilities into a common form, such as state money or gold, then each bank's money is interchangeable and hence indistinguishable. With convertibility, each bank individually cannot expand its lending, but the banking system together perhaps can (Keynes, 1930a, p. 26; Schumpeter, 1954, p. 730). This assertion ignores the level of reserves for the banking system as a whole, and the level of competitiveness within the system. Expansion is easier if the banks are not competing. Individual banks do compete by offering convertibility and the payment of interest on deposits, both of which place limits on the creation of bank money (Glasner, 1985, p. 58). Competition, and not reserve requirements, is the constraint on the banking system's issuing of money (Glasner, 1985, p. 57). In neoclassical terms, if each marginal investment or new loan is subject to diminishing returns, it follows that 'by paying competitive interest on deposits, each bank reaches a profit maximizing equilibrium at which the incentive, under given cost and demand conditions, to create additional notes and deposits is exhausted' (Glasner, 1989, p. 216; see also Tobin, 1963).

The efficacy of convertibility as a limit on banknote issuance is the crux of the historical debate between the 'Currency' and 'Banking' schools. For the Currency School, the value of paper money was derived solely from its correspondence to metallic currency, with the implication that regulations were required to limit its issue and prevent price inflation, a view derived from the quantity theory of money (Ingham, 2004a, pp. 41-42). For example, one fear was that paper money would mask the effects of an efflux of gold and prevent the money supply from contracting in accordance with Hume's (1742, Chapter II.V) price-specie-flow mechanism.[15] The Currency School recognised that gold and state money must be used in final settlement so that the growth of the bank credit superstructure could not get too far out of alignment. Since Bank Rate was used to protect the gold reserve, the hierarchy of money would impose itself, but with a problematic time lag and instability (Sayers, 1960, p. 718).

For the Banking School, supported by the 'Real Bills Doctrine' and the 'Law of Reflux', regulatory control of banknote issue was unnecessary. The Real Bills Doctrine—the

[15] For critiques of Hume's flawed price-specie-flow mechanism see Godley and Lavoie (2012, pp. 196-197) and Samuelson ([1980] 2015).

recommendation that bank-issued notes should only be backed by non-speculative or trade-based bills of exchange—would ensure that note issue (known as efflux) would only meet the 'needs of trade', thereby 'controlling the quality of money' (Mehrling, 1996, p. 332). The Law of Reflux states that note convertibility, by providing a means to return unwanted banknotes to the issuer in return for deposits, which must pay interest, makes an over-issue impossible. An over-issue of notes was impossible and the convertibility into gold would be enough to ensure uniform world prices by the law of reflux (Glasner, 1985, p. 60).

Both schools agreed on the need for a gold standard and that the potential for an over-issue of notes existed. The Banking School relied on convertibility and the fact that deposits were no different from notes, whereas the Currency School merely wanted more control over Bank of England notes (Schumpeter, 1954, p. 728).

> The Banking School were right in emphasising the variety and the importance of sources of credit, but the Currency School were right in arguing that the superstructure of credit could not for very long get out of line with the supply of the basic money, the gold and bank-notes. (Sayers, 1960, p. 718)

The result of the debate was a modification of the Currency School's policy (Sayers, 1960, pp. 718-719) with the recommendation that the central bank must act as a lender of last resort (LOLR) to prevent instability (Bagehot, 1873). In the end, the differences between the Banking and Currency Schools were more of a practical nature than theoretical, but they serve to highlight the conflicting aspects and interpretations of the hierarchy of money.

That modern banks are legally required to convert deposits into state money on demand is an arbitrary institutional arrangement enforced by the state and is a prime factor that contributes to bank runs, credit crunches and insolvency.

> The abolition of the law of legal tender would help to mitigate such a situation [credit crunch or bank run] by making everybody realize that, once he had become a depositor in a bank, he had sold his credit to that bank and was not entitled to demand payment in coin or government obligations. (Innes, 1913, p. 405)

Just so, Galbraith (1975, p. 98) outlines historical periods where convertibility was suspended or not enforced. For instance, 'by the time of the [American] Civil War ... [a]n estimated 7000 different bank notes [inconvertible currency] were in greater or lesser degree of circulation, the issue of some 1600 different or defunct state banks.'

Convertibility is not an intrinsic characteristic of banking. Innes (1914, p. 153) outlines historical periods where bank money has been stronger than state money, thus showing that the order in the hierarchy of money is not immutable.

> It is only recently, perhaps since the consolidation of the British war debt, that state finance has been on a sufficiently sound basis to rival private finance. And it is even more recently that state finance has been on a sufficiently sound basis to back a convertible domestic currency. (Mehrling, 2000c, p. 7, n. 5)

Immediate convertibility is a consequence of making state money legal tender. Originally introduced to enforce the acceptability of the state monetary token (regardless of metal content) it has morphed into the only acceptable means of payment.

> The effect of this impression is peculiarly unfortunate. When suspicion arises in the minds of depositors, they immediately demand payment of their credit in coins or their equivalent namely a credit on the State bank, or "lawful money,"—a demand which cannot possibly be complied with, and the result is to augment the panic by the idea getting abroad that the bank is insolvent. Consequently at the beginning of a stringency, every bank tries to force its debtors to pay their debts in coin or credit on the government, and these debtors, in their turn, have to try to extract the same payment from their debtors, and to protect themselves, are thus forced to curtail their expenditure as much as possible. When this situation becomes general, buying and selling are restricted within comparatively narrow limits, and, as it is only by buying that credits can be reduced and by selling that debts can be paid, it comes to pass that everybody is clamoring for payment of the debts due to them and no one can pay them, because no one can sell. Thus the panic runs in a vicious circle. (Innes, 1913, p. 405).

That banks are required to convert their money into state money on demand has the effect of giving the impression that state money is the only money-proper. Hence, the current institutional setting of banks maintaining equality at par with central bank money is logically akin to the gold standard. As such, it could then be imagined that banks could move from this standard to an equivalent ‘fiat’ standard where they do not maintain convertibility into any commodity or money. With convertibility suspended or halted, bank monies would become distinguishable, with the result that a ‘depositor sells to his banker his right on someone else and, properly speaking, his sole right so long as the banker is

solvent, is to transfer his credit to someone else, should the latter choose to accept it' (Innes, 1913, 404). Instead, banks would issue their own individual money notes, which would be the only means possible for borrowers from that particular bank to repay their debts to that bank.

The acceptability of each bank's notes would depend on their 'monetary space' (Ingham, 2004a), that is, on the extent of each bank's reach into the activity of commerce in the region it serves. The value of each bank's notes would depend on the perceived creditworthiness of its customers. There would be variation in the value of bank moneys just as displayed by the US banking system in 1814 when 'the banks outside New England suspended specie payment' (Galbraith, 1975, p. 84). A hierarchy of banknote discounts developed:

> The notes of New England banks, since they were exchangeable into gold or silver, were accepted at par therewith. The slightly less promising notes of New York were subject to a discount of 10 per cent. The distinctly more garish notes of Baltimore and Washington banks had a 20 per cent discount. Numerous notes from west of the Appalachians were at a 50 per cent discount. (Galbraith, 1975, pp. 84-85)

In all likelihood, the outcome would be either a dramatic shrinking of the effective monetary sphere of any one bank or a consolidation of the banking industry as banks merged to provide the brand familiarity and diversification necessary to ensure that their own monetary liabilities maintained a level of stability suitable to their customers. Although each bank's liability would not be convertible into a common element, an interbank market and interbank reserves would probably still be necessary. As central banks that issue fiat currencies maintain reserves of other central bank currencies, banks with distinguishable monies would need to do the same.

Nevertheless, the current institutional arrangements include a commitment by the banking system to maintain immediate convertibility from its monetary liabilities into state money. The banking system has developed many techniques for managing this obligation.

4.4 Asset and Liability Management

Without the explicit LOLR support by a central bank, a commitment by the banking system to maintain convertibility to state or central bank money exposes each bank to liquidity

risk: the risk of a bank run (Goodhart, 2008). According to Hicks (1989, p. 59) the three forms of protection are: 1) the law of large numbers, but this could fall apart due to inter-dependencies, 2) a cash reserve and 3) the interbank market (liability management). Hicks, however, has missed another: 4) matched-book or term funding, which is a more traditional form of liability management. We first consider the nature of bank deposits before turning to the interbank market, cash reserves and matched-book funding.

'Deposits in commercial banks, whatever their origin, are constantly being used' (Dunbar, 1922, p. 27) so a bank must keep a stock of assets which provide 'easy conversion into cash in case of need' (Dunbar, 1922, p. 29). Traditionally, short-term lending is considered 'self-liquidating' (Sayers, 1964, p. 183) and is the preferred means of satisfying the near-term need for cash. Long-term assets pose a liquidity problem: a bank would need a portfolio of shiftable assets. Assets that meet this criterion might be government and near-government bonds, but traded equities would not. Contrary to modern practice, mortgages, now a standard bank asset, would be traditionally unacceptable due to their illiquidity (Dunbar, 1922, p. 29).

To manage the liquidity risk resulting from this mixture of liquid and illiquid assets two types of bank liabilities arise: demand or cash deposits; and term, savings or time deposits. The former are held 'for the purpose of making payment' and are therefore monetary, whereas the latter are a 'means of employing savings' (Keynes, 1930a, p. 36). Term deposits must first be converted into demand deposits before they can be used to make payment. That monetary liabilities can be converted into non-monetary liabilities is an often-overlooked channel for reflux, and the distinction between these two types of bank liabilities must be emphasised. The problem is compounded by including non-monetary deposit liabilities in many of the broader measures of the money supply. 'This duality of function is the clue to many difficulties in the modern Theory of Money and Credit and the source of some series confusions of thought' (Keynes, 1930b, p. 214). The liabilities of the banking system bifurcate by function:

> The first function of the banking system is to provide *payment transmission facilities* which typically consist of chequeable deposit liabilities ("transactions money"), thereby acting as a clearing house in transferring current payments by means of book entries. Their second function is *intermediation*, thereby providing "asset money" (or rather indirect monetary and non-monetary liabilities in

general). (Bibow, 1995, p. 660, original emphasis)

The amount of each functional type depends on the term of the lending by the bank; the interplay between the tenor of a bank's liability structure and the convertibility of its assets is all important. For instance, if term deposit rates are too low—in an attempt to increase profits—then banks will have too many monetary liabilities and, for any given set of assets, their 'balance-sheet liquidity' will deteriorate (Hicks, 1962).

The liquidity of a bank balance sheet depends on a comparison of the term of its deposits and its liquid assets (Brunnermeier, Gorton, & Krishnamurthy, 2011), where the liquid quality of assets will 'vary over time and over state of mind' (Hicks, 1989, p. 63). The purpose of a fixed-term deposit is twofold. First, the holder is unable to alter its structure. A term deposit has a definite maturity date and, until that time, it must maintain its current form, even if transferred to another holder. Second, a term deposit cannot be used to repay a debt to the issuing bank. As emphasised so far, money must be undated, and available to pay debts immediately.

> Debts due at a certain moment can only be cancelled by being offset against credits which become available at that moment; that is to say that a creditor cannot be compelled to accept in payment of a debt due to him an acknowledgment of indebtedness which he himself has given and which only falls due at a later time. (Innes, 1913, pp. 393-394)

In the sense that money can be applied to the repayment of debts, fixed-term bank liabilities are not money.

Two examples are helpful: term deposits and negotiable certificates of deposit (NCDs).[16] Each is structured so that the issuing bank has no obligation to provide convertibility. Each helps preserve the bank's balance-sheet liquidity. Both pay a fixed rate of interest at maturity. To preserve the bank's balance-sheet liquidity, the liquidity risk has been passed to the deposit holder. For instance, the term deposit holder has accepted a form of market illiquidity. A term deposit provides price stability at maturity, but cannot be broken without violating its terms, nor can it be transferred to another depositor. It has limited price stability but is otherwise entirely illiquid. The NCD, in comparison, is transferable in a secondary

[16] Chapter 5 provides a more detailed discussion of the various forms of liquidity referred to here.

market and is liquid in a marketability sense, but, because a fixed rate of interest is paid at maturity, the value of the certificate will vary over time, exposing the holder to market risk. Again, the bank has protected its liquidity position by placing market-liquidity risk onto the NCD holder.

As Minsky (1957) points out in his discussion of financial innovation, profit-seeking activities have the capacity to structurally change the financial system, especially during periods of official monetary restraint, and alter the efficacy of monetary policy and the behaviour of the financial system in ways that are not fully understood at the time. One innovation that minimises the use of central bank money is the interbank market.

4.5 Interbank Market

In a banking system with many banks, banks ultimately use ‘bankers' money’ to settle interbank payments; examples are gold, central bank reserves, and deposits at correspondent banks (Minsky, [1986] 2008, pp. 225-226). In the simplest case—when two parties conducting a transaction have accounts at a single bank—payments can be effected merely by means of transferring ownership of the bank’s liability. If each party has accounts at different banks, then the transaction will enter the pool of inevitable two-way transactions between the two banks, and the amount owing from one bank to another can simply be settled as a deposit by the creditor bank with the debtor bank. The interbank market yields the simple outcome that the receiver’s bank replaces the original depositor, and there is no requirement for a transfer of central bank money. In effect, the receiving bank lends the paying bank the funds to make the transaction. That some banks are habitually lending banks and others are deposit taking or savings banks means that this is standard practice.

This process is simplified even further by means of a clearing-house system whereby the bilateral process of transaction netting is made with a single central counterparty, and only the net amount of transactions is settled (Dunbar, 1922, p. 47). In the course of many transactions, each bank will attempt to net out transaction traffic so that only net deposit balances between banks are established. Effectively ‘banks use the liabilities of the government only for net clearing’ (Tymoigne & Wray, 2006, p.13).

Under a net settlement system, there is no requirement for central bank balances in the normal course of events. The banking system can provide payment services almost without regard to the existence of state money (Turner, 2013, p. 3). Since ‘a debt can be just as well

cancelled against another debt as extinguished by a payment of money', in world without money, a banker 'can obtain credit from another banker' (Hawtrey, 1919, p. 4). The introduction of real-time gross settlement (RTGS), however, has increased the need for central bank funds for intraday interbank payments. The central bank must supply collateralised intraday loans (by means of repo) to ensure that each bank has sufficient central bank funds to handle these payments. These loans are repaid at the end of the day, so overnight balances at the central bank are small (Dent & Dison, 2012). As the monopoly supplier of the RTGS funds, the central bank is able to enforce its policy rate because 'it is always in a position to impose losses on individual banks that do no manage their cash flows effectively or on the banking system as a whole when it seeks to raise its target rate' (Rogers, 2008a, p. 23).

As stated in Section 4.3, when banks maintain convertibility, they operate as if on a gold standard, where instead of gold they agree to exchange their liabilities for central bank money at par. Each bank within a state monetary sphere needs to ensure that its balance of payments, or transfers of central bank money, is sustainable. Their means of ensuring an appropriate balance of payments is the interest rate that they pay on their liabilities: in the face of a central bank money outflow, they must raise rates, and lower them if there is too much inflow. This *modus operandi* is exactly that of central banks on a gold standard and means that banks must pay market rates on deposits.

Just as many central banks would maintain parity with gold, many commercial banks adopt a link with one central bank. The interbank market provides the means for banks to lend each other the central bank money required to satisfy payment imbalances in the same way as, under the gold standard, central banks would lend each other gold reserves in times of payment imbalances to maintain an orderly currency market (De Cecco, 1991, p. 321). This ability to lend the means of payment between banks has implications for the dynamics of the banking system.

4.6 Central Bank Reserves

That a deposit transferred from one bank to another is ultimately settled by means of a transferral of a deposit, or bankers' money, at the central bank has often led to the conclusion that central bank reserve balances play a significant role in the lending decisions of banks. The thinking is that, if an individual bank knows that new lending will require

the surrendering of its store of reserves to other banks, it will be hesitant in its lending. By this reasoning, if a bank deems it prudent to keep a 'fractional reserve' equivalent to 10% of its deposit liabilities then $100 of central bank money can only be scaled up to $900 of bank money in a process known as the 'money multiplier' (Ferguson, 2008, p. 50; Ingham, 2004a, p. 139). This process, however, does not distinguish between on-demand and term deposits or take account of the influence of interest rates. The theory of the money multiplier is supported by the observation that in many jurisdictions banks are required to maintain a reserve of central bank money as a proportion of their deposit liabilities.

Cash reserves are conceptually no different from gold reserves for a central bank operating under a gold standard and are ultimately managed by adjustments to the bank's discount rate. For commercial banks, the story is the same: term deposits are raised to ensure that a bank's liquidity mismatch is mitigated while still maintaining the desired level of profitability. Term funding enables a bank to match the maturity profile of its portfolio of bills or other loans. Banks, at an individual level, must postpone the possibility of reflux or convertibility for some portion of their liabilities to maintain their cash reserve, and this activity affects the liability structure of the banking system as whole.

The operation of monetary policy by a 'channel system' reveals the irrelevance of a cash reserve, and its independence from interest-rate policy (Lavoie, 2010).[17] Under this system, the central bank lends settlement funds throughout the day to those banks needing to make interbank payments. At the end of the trading day, some banks will have central bank deposits and some will be indebted to the central bank. All banks are expected to repay outstanding balances by the day's end. If a bank ends the day's trading owing money to the central bank, it must borrow the funds to repay the central bank from another bank—an exactly offsetting balance must exist in aggregate somewhere in the system, since other banks will have ended the day in surplus. At the end of the day the central bank ends up entirely square, with no outstanding balances of central bank money (Lavoie, 2006, pp. 19-21).[18] Although the rate of interest is indeterminate and market-driven, in practice it

[17] Examples of central banks operating one form of channel system or another are the Bank of Canada, the Reserve Bank of Australia, and the Reserve Bank of New Zealand (Woodford, 2003, p. 26).

[18] The Reserve Bank of Australia requires banks to hold a positive balance of Exchange Settlement funds (Reserve Bank of Australia, 2016). The Bank of Canada aims for a zero position (Lavoie, 2006, p. 20).

establishes itself exactly at the central bank's policy level.

A central bank operating a channel system is not an active participant in the overnight money market. Officially, the central bank lends at a rate slightly above its overnight policy rate (generally 25 basis points) and pays interest on deposits at a rate slightly below (again, generally 25 basis points). The central bank stands ready to borrow or lend at rates that provide boundaries to their guidance or policy rate, so that the central bank's policy rate operates in a 'channel' between its borrowing and lending rates. Any intraday loan of reserves taken by a bank in payment to another bank is borrowed back and the loan to the central bank is repaid. What remains is the bilateral loan between the two banks: the surplus bank has lent the deficit bank the funds to repay the original loan from the central bank; the central bank provides temporary elastic funds only. The result is that gross balances in promises to pay central bank money are high, but the net balance at the central bank is zero. These 'procedures ensure a determinate demand for net settlement balances – equal to zero' (Lavoie, 2006, p. 20).

Since the target balances for overnight central bank are zero, the central bank does not, in practice, make any payments based on its deposit rate or receive any funds at its lending rate unless the market is malfunctioning. The balance at the central bank will only be non-zero if borrowers and lenders in the interbank market experience difficulties trading with each other, such as when the payment flows are suitably unusual that surplus and deficit banks are not in contact, or when surplus banks develop an aversion to lending in the interbank market. In the case of a malfunctioning market, the central bank will need to act as an intermediary, lending to the deficit banks and taking the deposits of the surplus banks itself. In this case, the central bank will earn profits based on the spread between its lending and deposit rates (generally 50 basis points).

Similarly, if the quantity theoretic explanation of the value of money is taken to be in terms of central bank money in the equation of exchange then, in static equilibrium, the velocity of a stock of central bank money that approaches zero is effectively infinite. By extension, when a requirement to hold a reserve deposit at the central bank is imposed it can be seen that is has no actual effect on interbank clearing—it must remain fixed as a reserve at all times. Since central banks generally pay no interest on required reserves, they are simply a tax on the banking system in that they are loss-making interest-free loans to the central bank (Fama, 1980, p. 47).

The overall result is that, in general, banks use promises to pay in the unit of account for payment.[19] During the process of net clearing, the banking system will establish a zero position in central bank money by borrowing and lending between the deficit and surplus banks. The decision to lend money in the interbank market is to release control over it; the interest paid overnight is not paid on central bank, or 'base money', as assumed in many models, especially the New Keynesian variety (for instance, see Woodford, 2003, p. 75). It is interest paid on interbank credit. The fact that interbank borrowing and lending are done at the guidance rate of the central bank should not obscure the fact that interest paid on money in the channel system is not a 'risk-free' rate paid by the central bank. Instead, it is an (often unsecured) interbank rate made on promises to pay central bank money, or, more generally, credit measured in the unit of account. The implications of central banks that operate a channel system paying interest on reserves are less significant when it is recognised that the central banks also maintain zero, or near-zero, overnight balances. Consequently, the lending decisions of banks are not based on the reserves of central bank money available 'for lending' as in the money multiplier model (Jakab & Kumhof, 2015; McLeay, Radia, & Thomas, 2014).

As a clarification, Mehrling (2000a, p. 404) invites us to consider bank money as a derivative security over central bank money, since conceptually 'bank deposits are long positions in fiat money, while loans are shorts, and the outstanding stock of fiat money is the inventory.'[20] Like any other financial derivative[21], the stock of bank money at any one

---

[19] The nature of interbank market has changed significantly in the US since the GFC. Banks no longer undertake unsecured interbank lending and Fed reserves are the predominant settlement medium (Pozsar, 2017).

[20] The use of the long/short terminology here is somewhat confusing. In finance, a long position is one where the asset in question is held for a period of time; and in a short position the asset is sold and repurchased later. In each case there is an equal and opposite monetary transfer. In the case of long positions in money itself, it is ambiguous what has been transferred in return for the money held. In fact, in the case of a deposit the bank has issued its own money in return for the long position in state money. Nevertheless, the idea that a bank has adopted a derivative position in another financial asset is valuable in considering the nature of bank money and its relation to central bank money.

[21] A financial derivative is contract between two parties that derives its value from the price of another

time can dramatically exceed the stock of the underlying security, in this case fiat money. 'Viewing the bank money supply as the open interest in currency tends to shift our attention away from reserve requirements and money multipliers, and forces us instead to confront head-on the credit character of modern money, and its consequent elasticity' (Mehrling, 2000a, p. 405). Finally, the banks should not just be seen as just intermediaries, but as actually taking the long and short positions themselves via their deposits and loans, respectively. The role of bank lending, and the lender of last resort, 'enters naturally' (Mehrling, 2000a, p. 405), since, by taking long and short positions that are not perfectly offsetting, the bank is taking risk. When no central bank exists such as in the Eurodollar market, this risk must be managed more actively and explicitly.

4.7 Matched-Book Funding

The offshore Eurodollar market, which is the 'central institution of the world liquidity system' (Mehrling, 2015, p. 315), operates without the direct backstop of a central bank, since the Eurodollar is 'a private liability of global banks not the public liability of a central bank' (Mehrling, 2015, p. 320). Consequently, the market must pay particular attention to the timing of payments. Payment times are aligned by the use of foreign-exchange forward contracts to hedge unwanted risk and these contracts are supported by speculators who accept the risk. The price of accepting this risk determines the forward exchange rate, so that the forward rate is not an unbiased predictor of the expected future spot rate as per the expectations theory. This deviation is the expected source of profits for these speculators, so that 'uncovered interest parity...is incompatible with a system of private liquidity provision' (Mehrling, 2015, p. 316).

As outlined in Mehrling (2013b) forward exchanges rates are, in practice, set by foreign exchange dealers to match their order flow, which, since they are absorbing liquidity mismatches, must be maintained as inventory on their balance sheets until they can be cleared. As their positions become more extreme, that is they take more liquidity risk, the prices with which they are willing to trade move further from those they would quote in a 'neutral position' (Mehrling, 2013b, p. 358). As a result, the difference between the forward

---

underlying asset. Without an obligation for either party to deliver the underlying asset, no practical restriction to the creation of such contracts exists.

exchange rate and the current spot exchange rate becomes more pronounced. The term (not overnight) foreign and domestic interest rates, which are derived by arbitrage from the covered interest parity, are affected as well. Interest rates provide the adjustment mechanism for managing liquidity risk.

Balance-sheet size restrictions introduced by the Basel III regulations and money market fund reforms are limiting the ability of dealers to absorb this mismatch in order flow (Pozsar, 2016). This lowering of dealer leverage has caused a violation of the covered interest-rate parity (CIP) arbitrage relationship and a persistent cross-currency swap basis (Sushko, Borio, McCauley, & McGuire, 2016).[22] Balance-sheet restrictions mean a failure to exploit this arbitrage, with the result that liquidity is currently being priced in the cross-currency swaps market (Mehrling, 2015). The 'deviation from CIP is quite a clean measure of the price placed by banks on the use of their balance sheet' (Shin, 2016, p. 7) or the 'shadow price of bank balance sheet capacity' (Shin, 2016, p. 12). Liquidity is a balance-sheet phenomenon provided by dealers and banks.

4.8 Clearing and Final Settlement

The distinction between money-proper and credit money is generally based on the ability of state money, as money-proper, to perform 'final settlement' or credit clearing, whereas credit money is postponed settlement. In this view, debts are only settled when they are repaid by state money. Under the credit view, however, state money is simply credit and hence this form of settlement is not sufficiently final. Under a stricter definition of final settlement, all debt, and hence money, is extinguished. Payment by currency, or state money, is not enough:

> The payment of clearing house balances in this way [use of state money for settlement] could not occur unless the currency were redundant. It is not really payment at all, it is a purely fictitious operation, the substitution of a debt due by the government for a debt due by a bank. Payment involves complete cancellation

[22] By CIP, the forward exchange rate, F, is related to the current spot exchange rate, S, by the relationship $F = S(1 + i_f)/(1 + i_d)$, where i_f and i_d are the respective foreign and domestic interest rates. Deviations from this relationship imply an extra interest rate spread, or cross-currency basis, which, as a source of riskless profit, should be eliminated. Its persistence is evidence of the balance-sheet limits to arbitrage.

> of two debts and two credits, and this cancellation is the only legitimate way of paying clearing house debts. (Innes, 1914, p. 167)

The definition of payment implied here is much more stringent than many would impose. By this definition, the settling of interbank payments with promises to pay is not payment at all. Similarly, a payment by the transfer of central bank money is not full payment either. Conceptually, payment with state money is merely a swap of one debt for another debt. To see this, consider that banks will accept central money for settlement of interbank domestic transactions, but central bank money is not necessarily acceptable for international transactions (Michell, 2017).

Instead, complete and final payment requires that all balance-sheet relationships be cancelled, a phenomenon that can only occur when each of the two agents involved owe and are owed the same amount to each other. No state money is required for this form of payment. For example, banks accept their own liabilities as payment for the debts that private agents incur with them. Banks do not accept non-bank private credits. The repayment of a loan from a bank requires the re-acquisition of a bank deposit of the equivalent amount to the loan outstanding. Then all four balance-sheet entries between the borrower and the bank are cancelled. Bank loans are not repaid with state money, they are repaid with bank money, the distinction between money-proper and credit is misleading. *In extremis,* if all payments were satisfied to this degree, then no money stock would exist.

Hence it is instructive to reemphasise that money is an undated credit ‘redeemable at any time’ (Tymoigne, 2017, p. 1). In effect, money has a zero-length maturity. We can use the idea of a financial put option, which gives the holder the right, but not the obligation, to sell a specified asset to the put option writer at a pre-agreed strike or exercise price, to interpret money. In finance terms, money is a zero-coupon perpetual bond with an embedded put option to the benefit of the owner, where the asset is the perpetual bond and the exercise price is par.[23] Reflux is the simply the action of exercising this put option, whereby the holder of money repays taxes or a loan from a bank. In recognising this put option, it is tempting to use it to derive a floor price of 100% of nominal money value, but this approach

[23] I have drawn on discussions in Romanchuk (2016) and Rowe (2016).

would be ignoring that all financial instruments routinely have a value significantly different from their nominal value due to factors such as uncertainty, time, credit risk, to name only a few.

The put option is of the nature of an 'American' option[24], meaning that it can be exercised at any time. It is a standard maxim in finance theory that an American option should never be exercised early; it always has a higher exchange value than exercise value before maturity. The same maxim can be applied to the put option embedded in money to explain the demand for money. Furthermore, the volatility of the underlying asset of an option contract affects the value of the option itself. If we consider fundamental uncertainty as the 'underlying' of money then a direct conceptual link between uncertainty and a demand for money can be established.

The value of money is sustained in three ways: first, by reflux to the issuer, either as a repayment or converted to another debt form without the put option; second, by transferral to another agent who wishes to perform this reflux; or, third, to be held by an agent for its liquidity value, that is, for the latent benefit of its put option.

4.9 Banking Theory

The ontological picture of the banking system established in this chapter allows us to consider the three predominant theories of banking behaviour and the provision of money: debt intermediation, fractional reserve and credit creation (Werner, 2015).[25] The first, the debt intermediation theory, is most closely aligned to the cloakroom ticket view of banking whereby banks simply re-lend money placed on deposit. It disregards the fundamental banking activity as a swapping of IOUs, where a bank's IOU is monetary, and assumes that the banking system simply exists to channel the flow of savings into investment opportunities: Money is neutral and banks have no specific macroeconomic implications. Fractional reserve banking, on the other hand, recognises bank money and focuses on the limitations of deposit creation as a behavioural multiple of the amount of state-money reserves available to the banking system. This money multiplier fixes the amount of bank

[24] By contrast, a 'European' option can only be exercised at a fixed maturity date.

[25] Chick (1992) discusses the evolution of the banking system where the stages of development overlap with the three theories (see Section 9.4)

money in terms of the amount of state money, so that the banking system takes no part in macroeconomic analysis. Finally, the credit-creation theory works from the observation that bank lending involves a balance-sheet expansion, whereby a loan asset and deposit liability are created simultaneously as loan commitments are drawn by borrowers. The essential feature of the credit-creation theory is that banks do not first need reserves in order to lend. The process of lending itself creates a deposit by an individual bank, not at the level of the banking system as in the fractional reserve theory.

Each theory captures a behavioural element of the banking system at the expense of other aspects. The debt intermediation theory is based on the observation that banks intermediate between borrowers and lenders but misses the process of lending which leads to the outcome. It assumes that banks must lend physical state money already on hand, and the savings provided to investment by the banking system must have already been generated. Rather than lending savings, banks lend purchasing power and confusion concerning this distinction arises from a lack of clarity in the meaning behind such terms as ‘money’, ‘funds’ and ‘saving’. The outcome of new lending, where agents receive a counterpart of a new deposit, is what Moore (1988, 2006) calls ‘non-volitional saving’ or ‘convenience lending’ to the banking system. Agents receive and hold bank deposits, not because of any prior saving decision, but as part of commercial activity. A subsequent decision involves portfolio re-allocation and the potential reflux of the deposit. Instead, the debt intermediation theory is concerned with the final state of equilibrium when all depositors have been converted by their own decisions into savers. The process of arriving at equilibrium and the dynamics of the flow of funds are not considered; ‘the neglect of the role of financial intermediation in current macroeconomic analysis with its characteristic preoccupation with equilibrium states restricts the explanatory power of macroeconomic theory immensely’ (Bibow, 1995, p. 664).

Commercial banks operate on a state-money standard, whereby their liabilities are convertible into state money on demand, an imposition that is reflected in the fractional reserve theory. A single bank alone cannot expand the bank-money supply because requests for conversion into state money will undermine it individually. Instead, each bank must manage the demand for convertibility by means of the interest it pays on its term, or non-monetary, deposit liabilities. This form of reflux, whereby monetary liabilities are converted into non-monetary liabilities, is not entirely incorporated into the fractional

reserve story. The fractional reserve theory provides a behavioural link between the state-money supply and the bank-money supply, so that the latter can be excluded from the quantity theory of money. As Lavoie (1984, pp. 792-793, n. 11) points out, however, a stable money-multiplier is not compatible with bank's maximising their profits by portfolio adjustments, since it implies a fluctuating credit multiplier instead of the stable one as seen in practice, resulting from the accommodating behaviour of central banks to ensure financial stability (Lavoie, 1984, p. 779).

The credit-creation theory is sound in that banks create deposits as the first by-product of lending. The existence of overdraft commitments whereby banks provide the public with ability to spend on demand is enough to establish this fact. It should also be recognised that, subsequently, the bank must honour its convertibility obligations by settling any deposit transfer to another banking entity by providing interbank or state money. Each bank's lending is limited by its ability to access interbank or deposit funding. If all banks within the system are expanding at a similar rate then the state money is available from within the system. The banking system's reserves of state money are not funds available for lending—they are already lent to the state. Any use of state money to settle interbank payments is not part of any credit transaction; it is merely a settlement of a credit transaction. The availability of state money, from the central bank or other banks, for final settlement directly affects interest rates, but only after the loan has been made.

Being purely descriptive, the credit-creation thesis falls short of being a theory. Merely observing that banks create a liability when they lend does not provide an explanation of this phenomenon, why it is possible, what conditions are necessary, and how much credit money can be created. Arguably, the fractional reserve theory answers some of these questions in equilibrium, but at the expense of ignoring the path to equilibrium. In the current institutional banking framework, bank money is convertible to, and pegged to the value of, state money. Therefore, convertible bank money cannot be used in the equation of exchange as the correct metric in the quantity theory of money, as does Werner (2015), since the quantity theory is only applicable (if at all) to non-convertible, non-commodity money.

4.10 Summary

Banks are clearing houses of credit. The banking system, once the credit view of money is

adopted, is an integral part of the monetary nexus. The dominance of state money in the prevailing literature can be seen as a form of myopia, while the real monetary action is conducted elsewhere. The adoption of the conceptual essence that banking is a swap of IOUs, coupled with a credit theory of money, releases the analysis from whether money is provided publicly or privately. The answer is both and each form has its strengths and weaknesses.

Money is a debt that is immediately payable to the issuer in return for the asset that backs it. In the case of the gold standard, the asset is gold; in the case of fiat money, it is taxes. In the case of bank money, it is the repayment of an outstanding loan. The balance between debts payable and credits outstanding in the current period determines its value. Without enough tax payable, state money loses its purchasing power. Similarly, banks must have credit assets repayable in the near future to support their monetary liabilities. The historical distinction between banks, which lent for trade, and building societies, which facilitated mortgage lending, displays early efforts to manage these mismatches in the timing of repayments.

Historically, the practical consequence for the banking system to maintain convertibility to state money are reserves of central bank money. More recently, liability management and central bank support as the lender of last resort have made these reserves unnecessary. Banks have developed a myriad of techniques both to manage the liquidity risk inherent in their activities and to minimise their dependence on state money. Central banking operating models such as the channel interest rate system demonstrate the deficiency of standard monetary theories such as the quantity theory of money, and closer inspection shows that modern interpretations are also found wanting.

The current feature that banks must convert their liabilities into state money is not historically or logically essential. Bank money can conceptually exist parallel to state money. An analysis of the banking system should recognise both its dependence on, and its independence of, the state. Banks are specialists in issuing and clearing credit and are able to create the means of payment and provide purchasing power on demand. Furthermore, a revised definition of final settlement, where balance-sheet entries must cancel, shows that state money is not logically necessary for repayment. Money is not conceptually a persistent stock, but a perishable flow continually created and destroyed. The ability of banks to manage their monetary versus their non-monetary liabilities is essential to this process and,

consequently, to the provision of liquidity.

The parallels between the modern banking system and the gold standard are striking. They highlight the inconsistencies in the current monetary literature, which habitually leaves bank-created money unconsidered. The gold standard provides better insights than the money multiplier. Commercial banks must maintain convertibility between their money and state money at a price of par, just as central banks used to convert their money into gold at a fixed weight. Commercial banks keep reserves of central bank money to maintain this link but the main tool they use is interest rates. Raising rates paid on deposits limits the flow of conversions. Thus, the creation of too much deposit money would require higher term-deposit rates to prevent the conversion of deposits into state money.

The importance of the banking system to the provision of money must be stressed, allowing state money to be put into the broader context and given its appropriate emphasis. The banking system specifically, and credit more generally, has the ability to facilitate trade without necessarily relying on state money. Credit clearing, which includes the complete cancellation of two sets of balance-sheet debits and credits entries, is the fundamental means of final payment or settlement.

In the next chapter, the ontological landscape is expanded to consider the nature of liquidity in more depth.

5 A Taxonomy of Liquidity

5.1 Introduction

The terms ‘liquid’ and ‘liquidity’ cover many loosely defined concepts. Liquidity is most commonly identified with being ‘more certainly realisable at short notice without loss’ (Keynes, 1930b, p. 67) or ‘the ability to convert an asset into money at short notice with minimum loss’ (Rogers, 2014, p. 5). Other equivalent terms such as ‘saleableness’ (Menger, 1892, p. 242), ‘shiftability’, ‘salability’ (Mehrling, 2011, p. 6) or ‘marketable’ (Kiyotaki & Wright, 1989, p. 935) are used to encompass the three dimensions: trading a significant amount at the asset’s fundamental price in a short time-frame. The fact that some assets are seemingly more marketable than others has led to the concept and empirical observation of a liquidity premium. According to Mehrling (2011, p. 50), ‘the extra yield on the long-term investment is a kind of “liquidity premium” that compensates the long-term investor for the fact that he may have to take a loss if for some reason he needs to convert his investment into cash before maturity.’

The liquidity literature classifies the common, saleability form of liquidity as ‘market liquidity’ and distinguishes it from ‘funding liquidity’, which refers to the ability to refinance debts as they mature (Brunnermeier & Pedersen, 2009). A third form, ‘technological liquidity’, arises when an investment in physical capital is reversible (Brunnermeier, Eisenbach, & Sannikov, 2013, p. 5). Market liquidity and technological liquidity are both related to the ease with which an investment can be reversed if necessary, since marketable financial assets allow reversible commitments to otherwise illiquid and heterogeneous capital assets (Brown, 2003). Market liquidity cannot apply to the economy as a whole in the same way that technological liquidity can. To a certain extent, market liquidity can mitigate technological illiquidity, although not for society as a whole.[26]

None of these concepts, however, covers the pure or perfect liquidity of money itself. Lozano (2015, p. 14, original emphasis) argues for a ‘universal definition’ of liquidity ‘*as*

[26] Non-specialist equipment with many applications is the sort of physical equipment that could be classified as technologically liquid. For instance, in economic models containing a single consumption good, such as corn, whereby saving and investment are indistinguishable, ‘investment’ can be reversed by consuming more corn. Technological liquidity is not related to conversion into money and will not be considered in this study.

the nominal relationship between maturity and value.' The term 'liquidity' has a metaphorical aspect that must be considered in relation to the real phenomena it represents (King, 2012). The use of the term 'liquidity' is a form of 'linguistic ingenuity' (Jespersen, 2009, p. 68) that introduces additional connotations, which influence how concepts are interpreted when insights from the analytical world are applied to the actual world of policy. This chapter examines in more detail the ontological aspects of the elements of the taxonomy of liquidity.

5.2 Market Liquidity

Market liquidity, whereby 'liquidity is the ease of trading a security' (Amihud, Mendelson, & Pedersen, 2005, p. 270), naturally leads to a study of the microstructure of financial markets. In an idealised world of perfect market liquidity, all assets can be immediately bought or sold in any amount at their fundamental value. Deviations from this perfect market liquidity arise from 'frictions' such as transaction fees and taxes, demand pressure, inventory risk, information asymmetries and search impediments. Each restriction adds costs to transactions and reduces the ease of trading, in other words: illiquidity. The costs of illiquidity can have an impact on asset prices and expected returns 'if investors require compensation for bearing them' (Amihud et al., 2005, p. 271). Market liquidity costs are paid either in 'shoe-leather' while searching or as income to securities dealers or market makers. For the market maker, the problem becomes one of inventory management:

> …variations in demand pressure that cause variations in the market maker inventory change the prices at which he is willing to trade. These are short-term, transitory effects of inventory on prices, but the permanent effect on prices and expected return flows through the effect on trading costs. For example, in market systems with better capacity to absorb inventory shocks, the models would predict smaller illiquidity costs and consequently there would be smaller price discount due to illiquidity. (Amihud et al., 2005, p. 300)

Like used-car dealers, market makers provide inventory management services for profit and are an important part of the ontological landscape of liquidity, just as bankers are for money. Market makers can profitably provide a ready market for people wanting to buy and sell by means of a spread between buy and sell prices. 'Dealers supply liquidity by absorbing temporary imbalances on their own balance sheets, and they charge for the

service' (Mehrling, 2013b, 359).

In providing their service, market makers hold assets in inventory until they can be resold, and by having a judicious combination of long and short positions they can largely avoid overall market or systematic risk. Dealers operate in the practical realm of finance and use relative values (Cochrane, 2005b, p. xiv); dealer-based liquidity is founded on relative value not absolute value. A 'dealer has no need to assess the fundamental value of the security' (Mehrling, 2013b, p. 356). By applying matched-book or hedging techniques (Stigum, 1990, p. 433) dealers can inoculate themselves from both systematic risk and the need to know absolute prices. Instead, they rely on the frequency of transactions and a stable relative-price structure. In the face of constant selling pressure, they must stand aside. The market makers themselves provide market liquidity and, as such, it is a property superimposed on the asset by a third party, not a fundamental or intrinsic characteristic of the asset itself. In a dramatic movement of market level market making ceases and all assets become illiquid until the new level of the market is established.

Inventory imbalances are resolved, not by adjusting the bid-offer spread, but by skewing bid and offer prices (Treynor, 1987). When a market maker's inventory is too long, both the bid and offer prices are lowered relative to those displayed by other market makers. Conversely, when the inventory is too short, the relative prices displayed are raised. In this way, flow imbalances are shared between market makers. Market makers operate a smaller 'inside' spread within the protective boundaries of the 'outside' spread provided by 'value-based investors' (Treynor, 1987). 'Dealers are thus valuable to transactors in a hurry, because they greatly reduce the spreads encountered by those transactors. By doing so, they also greatly improve the liquidity of the markets in which they deal' (Treynor, 1987, p. 27). In addition to skewing their bid and offer prices, market makers rely on the outside spread to manage their inventory. At the limits of a dealer's inventory capacity, the market maker can deal with 'the only other transactor in the market who is motivated by price—the value-based investor' (Treynor, 1987, p. 27). The value-based investor acts as a 'market-maker of last resort' (Treynor, 1987, p. 28) or, equivalently, 'dealer of last resort' (Mehrling, 2011).

Trading itself incorporates information. If trading happens at a low frequency, then there will be more information between each trade and the market will be less continuous. Large orders result in discontinuous price jumps. 'The ability to handle large amounts of stock in

short periods of time without changing the price of the stock is not a characteristic of a liquid market' (Black, 1971, p. 31). The market maker should not resist large changes in price resulting from large orders, nor should he increase his spread, as this will make prices 'sticky' (Black, 1971, p. 32). If prices move with new information, they will be informationally 'efficient' and exhibit randomness, meaning that profits cannot be earned from statistical knowledge of past transactions (Black, 1971, p. 32; Fama, 1970, 1991). Therefore, large price movements in one direction, and not a series of small steps, are a sign of an efficient market. If market makers resist price moves, then those with special information will make money at the expense of those without.

Fischer Black gives the clearest set of criteria for market liquidity provided by an organised exchange:

> …the market for a stock is liquid if the following conditions hold: (1) There are always bid and asked prices for the investor who wants to buy or sell small amounts of stock immediately. (2) The difference between the bid and asked prices (the spread) is always small. (3) An investor who is buying or selling a large amount of stock, in the absence of special information, can expect to do so over a long period of time at a price not very different, on average, from the current market price. (4) An investor can buy or sell a large block of stock immediately, but at a premium or discount that depends on the size of the block. The larger the block, the larger the premium or discount. (Black, 1971, p. 30)

In a liquid market, small amounts should be tradable at any time, but it is unrealistic to expect to be able 'to buy or sell large blocks of stock in short periods of time without moving the price of the stock very much' (Black, 1971, p. 30; see also Kyle, 1985). An organised exchange has low costs and bid-offer spreads, continuous trading or immediate execution, and easy access (Black, 1971, p. 34). It does not display 'price continuity or "stability"' (Black, 1971, p. 35). An attempt to sell a large block of an individual security suggests the possession of private information affecting relative values and not a simple need to raise funds. If a genuine desire for cash is the motivation for selling, then the investor should be happy to sell smaller amounts of any combination of the stocks in their portfolio. This observation alone demonstrates that discounts are inevitable for large parcels (Black, 1971, p. 30).

An ideal market has at most one market maker, more than this is inefficient (Black, 1971,

p. 33). In practice, market making requires an established liquid point of convergence in the form of an asset that can be bought and sold to offset the risk of the other, less-liquid, assets held inventory. Thus, many market makers, operating by hedging against a common futures contract, emulate an ideal single market maker in the same way that the banking system aims to operate as one bank by means of an interbank money market.

Echoing the line of reasoning followed in Section 4.8, market makers can be conceived of as providing both a call and a put option to their customers. The customer has the right, but not the obligation, to buy at the ask price and the equivalent ability to sell at the bid price (Copeland & Galai, 1983, p. 1464). Bid-ask spreads, and hence the cost of market liquidity, must increase with risk and volatility (Amihud, 2002, p. 39) and observations of volatility and market liquidity for both bonds and stocks are highly correlated (Chordia, Sarkar, & Subrahmanyam, 2005).

With market liquidity provided by an organised exchange the asset is ‘readily shiftable’, since, in general, a market price exists at all times, but nevertheless the gain or loss on sale is unpredictable (Sayers, 1964, p. 178). The asset has the property of being convertible into money quickly but not necessarily without loss. The uncertainty concerning the proceeds of an asset sale associated solely with market liquidity excludes it from any idea of ‘perfect liquidity’, as with money itself; instead, it merely signifies a form of ‘shiftability’ (Sayers, 1964, p. 179). For this reason, Robinson (1951, p. 94) calls market liquidity the ‘convenience’ of an asset and explicitly draws attention to the crucial difference between market liquidity, an asset’s underlying value, and the idealised concept of ‘perfect liquidity’.

For example, government bonds held to maturity have a known capital-value outcome, but this certainty of outcome comes at the cost of illiquidity. Convenience is a concept quite apart from variability in an asset’s price and distinguishes it from ‘capital-uncertainty’ and ‘income-uncertainty’ (Robinson, 1951, p. 94). Bills and bonds have the same level of convenience but are less convenient than money (or bank deposits) and are ‘good’ in the sense that there is no credit risk (Robinson, 1951, p. 95). Bills differ from bonds in that they have no capital-uncertainty, but they do have income-uncertainty.

In the practical realm, Kay (2015) finds that market liquidity might help in providing a continuously tradable price, otherwise known as ‘price discovery’. He makes the point, however, that many long-term investors would be just as well served by an asset market

that traded once per week. Instead, the attempt to offer perfect market liquidity in investment products affects the composition of the underlying assets and is ultimately counter to the interests of investors.

5.3 Funding Liquidity

Funding liquidity represents the ease with which a borrower can obtain a loan or access money to repay an existing loan. Whereas market liquidity refers to conversion of assets into cash and is therefore associated with the asset side of the balance sheet, funding liquidity is the counterpart for the liability side of the balance sheet.[27] Funding liquidity represents the ability to access funds by issuing new liabilities, or the re-establishment of an existing liability.

Nevertheless, funding liquidity can be re-interpreted as a form of market liquidity. Conceptually, the borrower is 'selling' an asset in the form of a new loan to replace an old one, or equivalently re-selling an old one, since '"borrowing" is selling an asset, the asset being money forward' (Radcliffe Committee, 1959, p. 132). The difference is that for market liquidity the asset being sold has only the risks associated with the asset itself, the nature of the seller of the asset is not important. For funding liquidity, on the other hand, the risks for the purchaser predominantly relate to the credit-worthiness of the vendor. Otherwise, the risks involved—transaction costs and the expected interest rate payable—are simply the ones analysed in relation to market liquidity. Since banks are the predominant providers of loans, they are essentially dealers in funding liquidity. In the presence of organised loan markets (overdrafts and pre-arranged lines of credit), where transaction costs are minimal, the interest rate applicable at the time that the loan is drawn is the predominant cause of uncertainty. Again, it is uncertainty of the future rate of interest at the time that the cash flow arises that is the fundamental element to be considered.

Dealers are the link between money markets and financial markets because repo provides

[27] For Tirole (2011, p. 288, n. 3), however, the correspondence of market and funding liquidity with opposing sides of the balance sheet is not so clear-cut. The counter-example given is the securitisation of a bank's loan portfolio coupled with liquidity support, but it actually shows the combination and the interaction of market and funding liquidity rather than a refutation of their boundaries. The balance-sheet correspondence is a useful working definition.

the funding for their market making. Funding liquidity and market liquidity are interdependent, both are mutually supporting (Adrian & Shin, 2010). The ease of trading overall depends on the interplay between both market and funding liquidity. Because inventories are funded by repo, market liquidity is constrained by the dealer's capacity for inventory, which in turn relies on funding liquidity. Brunnermeier and Pedersen (2009) and Mehrling (2010) emphasise the inherent roles of both market and funding liquidity in financial instability. The relationship between market liquidity, funding liquidity and risk suggests a 'commonality in liquidity' because the 'shadow cost of capital is a driving state variable' (Brunnermeier & Pedersen, 2009, p. 2227).

The link between market liquidity and funding liquidity can be explained by changes in monetary policy (Chordia et al., 2005). To see this, consider 'maturity transformation', where illiquid long-term debts are funded by short-term money-like debt. Maturity transformation blurs the distinction between money and credit and moves the liquidity risk from the investor to the financial system. The banking system creates market liquidity by adopting funding liquidity risk. The interdependence between market and funding liquidity identified above comes about due to their being fundamentally the same thing—the ability to sell financial assets. They are provided by dealers and are not macroeconomic phenomena.

5.4 Trade Liquidity

Considering a debtor's potential sources of finance reveals yet another form of liquidity. According to Minsky ([1986] 2008, p. 260) there are three sources of finance: 'cash flows from operations, refinancing or rolling over debts, and selling assets or net borrowing.' The final two are covered by the aforementioned funding liquidity and market liquidity, respectively; the first is not. Furthermore, Minsky ([1986] 2008, p. 223) identifies three types of operational cash flow: income, balance sheet, and portfolio. Again, portfolio cash flow arising from the sale and purchase of assets is already represented by market liquidity. The process of production and sale yields cash flow, and hence liquidity (Rochon, 1997, p. 287). We must introduce the category of trade liquidity, which is associated with the circular flow of funds, the dynamic process of income becoming expenditure, and the continual creation and destruction of money.

Income cash flows—wages and salaries, both public and private, the payments

> from one stage of production and trade to another, and gross profits after taxes of business—result from the process of production. Money, in effect, goes around an income circuit; the income circuit as here defined includes all the payments for partially finished products sold by one firm to another. (Minsky, [1986] 2008, p. 223)

The process of production, especially in buoyant conditions, allows for a form of liquidity based on expectations of future trade. In boom conditions, 'industry is likely to be confident of replenishing its liquidity out of future profits' (Radcliffe Committee, 1959, p. 103). Trade liquidity is available 'even when bank credit is being contracted, so long as business expectations remain sanguine' (Radcliffe Committee, 1959, p. 103).

> For consumption is just as effective in liquidating the short-term finance as saving is. There is no difference between the two. If the entrepreneur gets wind of ex-ante consumption in the mind of the consumer, he is not only just as safe to get liquid and pay off his bank in due course as where there is ex-ante saving, but indeed much safer—for there is no risk that the consumption, when it matures, will take the form of an enhanced desire for cash. (Keynes, 1937c, pp. 667-668)

The second type of operational cash flow—balance sheet—is income generated by existing assets and 'the rapidity with which the wealth embodied in them can become "liquid", in the sense of producing output, the proceeds of which can be re-embodied if desired in quite a different form' (Keynes, 1936, p. 240). For the purposes of liquidity, balance-sheet cash flow can be considered as trade liquidity.

Thus, the supply of liquidity, in its broadest sense, is not based on the number of liquid instruments in existence at any point in time. It must include some element of its potential supply. The relationship between liquidity, credit and money means that

> …an individual's power to purchase, which is the objective element behind demand in terms of *numéraire*, is not fully represented by the amount of the credit instruments that are actually used in "payment" or even, so we should add, by the deposits, overdrafts, etc. against which checks are drawn, but by the total amount that an individual could command if he wanted to, i.e. the amount that is actually at his disposal in some measurable form plus something that might be called *potential* credit, which defies measurement, yet is a factor in any situation.

(Schumpeter, 1954, p. 718, n. 5, original emphasis)

Because trade liquidity is based on the anticipation of spending arising from the proceeds of production, it is a socially constructed property of the system arising from the circular flow of income and expenditure.

5.5 Self-Liquidating Assets

One form of cash flow in the balance-sheet category that requires its own distinction is the form of liquidity displayed by short-dated financial instruments. Short-term assets, such as bills of exchange, are 'self-liquidating paper' (Sayers, 1964, p. 183), since they are expected to produce cash flow in the near future. These assets are generally classified as 'near-money' and have properties that are described as self-liquidating, since the 'proceeds of a loan would be used to finance the acquisition of a specific stock of goods, and the sale of these goods—either with or without processing or transporting—was to yield the funds to repay the debt' (Minsky, [1986] 2008, p. 229). Thus, it is a form of liquidity closely related to, but once removed from, trade liquidity, since the cash flow generated by the asset is supported by the proceeds of an anticipated income cash flow.

In the context of banking specifically, the term 'liquidity' can be taken to mean the 'ability to satisfy demands for cash in exchange for deposits', for which a banker 'must maintain an adequate degree of liquidity in his assets' (Sayers, 1964, p. 177). For a bank, the perfectly liquid asset is cash, by which is meant either state or central bank money, but it is an 'idle asset' since it generally earns no interest. To meet the requirement to satisfy depositors' demands for cash, bank assets must be convertible into cash quickly and without loss. The latter requirement means that banks cannot rely on assets that only offer only market liquidity because of their capital-uncertainty. As we have seen, government bonds are often highly shiftable, but have capital-uncertainty unless held to maturity, thus rendering them effectively illiquid for the purposes of a bank. The capital-uncertainty associated with a fixed-term asset decreases as it approaches maturity. The Real Bills Doctrine (see Section 4.3) recommends that banks hold self-liquidating assets (Mehrling, 1996, pp. 335-336; Mehrling, 2011, p. 31). A bank with a portfolio consisting entirely of long-term assets will be reliant on market liquidity, in other words, on shiftability to a market maker, and will be forced to secure term funding to match the illiquidity of its assets (see Section 4.7).

5.6 Aggregate Liquidity

The taxonomy outlined so far is incomplete; the pure or perfect liquidity associated with money itself is left uncategorised. If liquidity can only be defined 'in terms of exchangeability for money' then money cannot have 'perfect liquidity', for to do so 'is to argue in a circle' (Hicks, 1989, p. 42). Defining money as the most liquid commodity via its 'exchangeability' is therefore tautological (Ingham, 2004a, p. 6). Going further, liquidity presents 'both a tautology and a paradox' (Lozano, 2015, p. 15), for both liquidity in its purest form and the realised value of an asset are represented by money. Liquidity cannot only be money, for then they could not be separate concepts, nor could non-monetary assets have the property of liquidity (Lozano, 2015, p. 16).

Money has no natural place in a liquidity theory based simply on the ease of conversion to money (such as market liquidity), even if money is in some way defined to be perfectly liquidity or '*par excellence* "liquid"' (Keynes, 1936, p. 234, original emphasis). Under the commodity view, money is the most saleable, and hence liquid, good where its acceptability derives from its liquidity, and its liquidity from its acceptability. For Chick (1983, p.306), 'the fact that argument is circular does not make it less true. The properties of money are self-reinforcing.' The liquidity of money itself is of a form that is not captured by mere saleability, unless it refers to the conversion money into assets, the reverse of the process of saleability. The liquidity of money is in a different category from that of market liquidity.

Some degree of clarity is provided by the terms 'inside' liquidity and 'outside' liquidity, which are used to refer to liquidity provided by a cash reserve and liquidity from sale of assets, respectively. These terms have been associated with funding and market liquidity (Bolton, Santos, & Scheinkman, 2011, p. 266). Although it would be accurate to recognise outside liquidity as a form of market liquidity, holding a cash reserve cannot be justifiably viewed as a form of funding liquidity since no requirement for refinancing exists. Instead, including money in the framework requires a distinction between, on the one hand, 'market microstructure or microeconomic liquidity' (Tirole, 2008, p. 54), which focuses on bid-ask spreads and market liquidity, and, on the other hand, 'aggregate or macroeconomic liquidity' (Tirole, 2008, p. 55), which attaches to securities that do not lose value in an economic crisis.

The importance of a macroeconomic perspective is highlighted by the 'paradox of

liquidity', in which the 'the attempt of economic agents to become more liquid transforms previously liquid assets into not-so-liquid assets' (Lavoie, 2014, p. 21). Market liquidity is subject to a fallacy of composition: the more people want market liquidity the more it disappears. More precisely, a desire for aggregate liquidity undermines the supply of market liquidity. Nesvetailova recognises and expands on the illusion of market liquidity:

> The first fallacy is the assumption that it is the market-making capacity of financial intermediaries to identify, price and trade new financial products that creates and distributes liquidity in the markets. Second is the view that general market trade and turnover are synonymous with market liquidity. The third and corresponding fallacy is the notion that market liquidity itself – when multiplied across many markets – ultimately is synonymous with the liquidity (and financial robustness) of the economic system as a whole. Altogether, this line of reasoning has been underpinned by the notion that financial innovation in its various forms ultimately enhances the liquidity of the financial system as a whole. (Nesvetailova, 2010, p. 17)

The recognition that dealers can provide an overlay of market liquidity to assets only at a microeconomic level leaves the open question of what 'liquidity' can mean at a macroeconomic or aggregate level, and whether it can be measured. Thus, in the taxonomy of liquidity, aggregate or macroeconomic liquidity is an important classification. To extract the precise liquidity implications of the fallacy of composition, we should introduce yet another classification: aggregate liquidity. Two conditions appear to distinguish aggregate liquidity from market liquidity. First, aggregate liquidity does not require market-making agents for its provision and, second, it is available regardless of the prevailing economic conditions. The nature of agents that have the ability to provide macroeconomic liquidity compared with microeconomic liquidity leads to the idea of 'public' and 'private' liquidity, respectively (Holmström & Tirole, 1998). This distinction suggests another paradox of liquidity: financial innovations that appear to increase the provision of private liquidity actually decrease the level of overall liquidity (Lavoie, 2014, p. 21; Nesvetailova, 2007, p. 78).

For Nesvetailova (2010, p. 8, original emphasis), aggregate 'liquidity is a property of an *asset*', not the state of the market overall. Any assumption otherwise is rejected as a fallacy:

> At the level of financial institutions themselves, the axiom that financial innovation

> and engineering have the capacity to *liquefy* any type of asset – or, more accurately, debt – has resulted in the now mainstream notion of liquidity that is divorced from any attribute of assets *per se*. (Nesvetailova, 2010, p. 16, original emphasis)

Some assets are more liquid, in the aggregate sense, than others. Furthermore, aggregate liquidity is 'intimately related to the notion of money' (Nesvetailova, 2010, p. 8). Aggregate liquidity manifests itself only in a select set of assets:

> "Aggregate or macroeconomic liquidity" understanding: according to this alternative definition, variants of which date back to Keynes and Hicks, an asset offers liquidity to the corporate world if it can be used by the latter as a cushion to address pressing needs. To be an effective cushion, though, the asset must not lose value in those very circumstances in which the corporate sector does need money. In this respect, the on-the-run Treasury bond is rather distinct from the stock index or the hypothetical mortgage-backed portfolio in that it does not lose value in recessions, while the latter's value is likely to be reduced precisely in case of an industrial or financial recession. (Tirole, 2008, p. 55)

The description given here, however, is insufficiently distinguished from a more standard measure of risk or volatility, especially in correlation with systematic risk. It is suggested here that on-the-run Treasuries[28] are more liquid in a macroeconomic sense because they are negatively correlated with overall market risk. In many ways, this definition leads us simply to the 'risk-free' asset so prevalent in much of finance and macroeconomic theory. Aggregate liquidity should not be confused with asset-price uncertainty.

For Nesvetailova the definition of liquidity is 'an asset's capability over time of being realised in the form of funds available for immediate consumption or reinvestment – proximately in the form of money' (Nesvetailova, 2010, p. 8). Her discussion frames the historical movement in the meaning of the term 'liquidity' from a property of an asset, to a property of the market overall. It is clear from the context that Nesvetailova is implicitly referring to a process whereby the meaning of the term 'liquidity' has transitioned from aggregate liquidity to market liquidity and is discussing the associated macroeconomic

[28] The on-the-run bond is the one most recently sold into the market by the US Treasury and is the one most actively traded. Off-the-run bonds are held as core portfolio positions by insurance and investment funds and are traded much less frequently.

difficulties. The result is that ‘over the past few decades, analyses of finance in the macro-economy have assumed that liquidity is no longer primarily a property of assets, but rather an indicator of the general condition and vitality of a financial market’ (Nesvetailova, 2010, p. 11). The lack of a clear taxonomy of liquidity contributes to this confusion.

Similar confusion arises when Holmström and Tirole (1998, p. 2) focus on the demand for liquidity, where ‘liquidity refers to the availability of instruments (market and nonmarket) that can be used to transfer wealth across periods.’ These intertemporally liquid instruments are required to cover unexpected funding shortfalls. In their model, when risks are purely idiosyncratic, ‘private liquidity’ is sufficient, but in the face of pure aggregate uncertainty, ‘public liquidity’ must be provided. At an aggregate level, shiftability can only be awarded to assets by the ultimate market maker, the central bank (Sayers, 1964, p. 181). Thus:

> An institution that performs a lender-of-last-resort function guarantees that the terms of some contracts will be fulfilled, regardless of market conditions or the business situation of the particular debtor. Thus, a lender of last resort diminishes the risk of default of the assets it guarantees. Assets with low default risk are readily marketable—they are liquid. When the Federal Reserve extends the domain of instruments that it protects against default, it is increasing the effective quantity of liquid assets and thus of assets that have the properties of money in the community. (Minsky, [1986] 2008, p. 47)

As we have seen, however, marketability is not risk free: the central bank does not necessarily remove price uncertainty. The ability to avoid price uncertainty is valuable and an important part of liquidity.

5.7 Price-Protection

Ricks (2011) argues that money-market instruments function as ‘money’, where the key aspect of these ‘money-claims’ is their short-term price stability. Although a precautionary motive (Keynes, 1936, p. 170) describes why agents hold a stock of the medium of exchange, Ricks argues that the actual medium itself need not be held. Instead a very liquid equivalent will do—hence the existence of the money market and its highly liquid instruments. These instruments are more than just marketable, like equities or Treasury bonds: ‘Not only are they extremely liquid (i.e., convertible into the medium of exchange on very short notice and at practically no cost), but, like deposits, their value in terms of

money is almost always extremely stable' (Ricks, 2011, p.92). Ricks calls this their 'price-protection' feature, and it is more than just high credit quality.

Closely related to price-protection is the concept of 'information insensitivity' (Cochrane, 2014; Dang, Gorton & Holmström, 2012), which refers to debt instruments of sufficient credit quality that almost full payoffs can be expected in bad states, thereby removing the need for investors to uncover any private information about the underlying obligor. Not all money-claims can be used as a medium of exchange (Michell, 2017), but their near perfect market liquidity is their distinguishing feature and inasmuch as the shadow banking was able to create these instruments then this appears to be liquidity creation.

Using the concept of price-protection, Ricks (2011, p. 101) considers whether the Modigliani-Miller theorem applies to banks and other maturity-transformation firms (such as shadow banks) who finance themselves in the money market. The Modigliani-Miller theorem (Modigliani & Miller, 1958) demonstrates the irrelevance of debt versus equity funding in determining a firm's value. Money-claim financing, however, has a lower expected return because of the instrumental value of transaction-reserve assets, and the ability to create these instruments affects the value of the firm. The Modigliani-Miller theorem rests on the assumption that investors can re-create the financing arrangements of the firm themselves and hence adjust it to their preference. This is not possible with money-market funding. Thus:

> On closer examination, this analysis does not actually contradict the Modigliani-Miller theorem. That theorem rests on explicit assumptions, one of which is that the firm's suppliers of funds are able to finance themselves on the same terms as the firm. And this condition cannot hold true in the case of maturity-transformation firms. The market will not fund just any economic agent with money-claims; the 'moneyness' of a money-claim depends on its price-protection, which requires exceptionally high credit quality. For this reason, maturity-transformation firms invariably invest the vast majority of their assets in credit instruments—that is, in senior claims on other economic agents. As a logical matter, not every economic agent can limit its assets to senior claims on other economic agents. (Nor should they want to!) Thus, very few savers seeking to store transaction reserves are able to issue low-yield money-claims themselves. Otherwise it would happen all the time. (Ricks, 2011, p. 102)

This insight points the way to one potential limit to liquidity creation: the availability of senior claims on other economic agents. For Ricks this reliance on senior claims is a modern version of the money-multiplier. 'Functionally speaking, all of these firms are engaged in the activity of fractional-reserve banking (or money creation), whether they are called "banks" or not' (Ricks, 2011, p. 97). Thus, for Ricks at least, the shadow banking edifice of liquidity is built on foundations that are not unlimited.

The concept of price-protection has a broader application. Assets that maintain price stability in the unit of account provide a reserve of liquidity. Through the maintenance of convertibility, bank liabilities are one such example. Price-protection can also take the form of guarantee or risk transfer:

> In the private market, the amount of cash you can get for an asset depends on that asset's current market value. By buying a guarantee of the market value of your assets, in effect you are guaranteeing your access to cash as needed; if no one else will give you cash for them, the guarantor will. (Mehrling et al., 2013, pp. 7-8)

The hierarchy of money can then be extended to include the liquidity-storage capability inherent in price-protected money-claims (Pozsar, 2014). These money-claims come in four forms depending on the 'type of assets backing them and the type of backstops supporting them' (Pozsar, 2015, p. 4). First are public money-claims such as central bank reserves and government bills. Second, we have private-public money-claims in the form of government-insured bank deposits. Third, repurchase agreements backed by government securities are public-private money-claims, where private promises to pay are backed by public assets. Finally, there are purely private money-claims not supported in any way by a public assets or backstops. The result is what Pozsar (2014, 2015) calls the 'Money Matrix' (see Table 1). As noted above, many of these 'money' forms cannot be used as a medium of exchange, so they are more appropriately classified as forms of liquidity. The instance of assets with a price-protection feature, but which lack monetary characteristics, in that they cannot be used to settle debts, is an important phenomenon in the taxonomy of liquidity.

Table 1 – The 'Money Matrix'

	Public Backstop	**Private Backstop**
Public Assets	Central bank reserves, treasury bills	Government bond repos
Private Assets	Insured bank deposits	Non-government bond repos, uninsured bank deposits

(Adapted from Pozsar, 2014, 2015)

5.8 Summary

Market liquidity—the ability to sell assets quickly without undue loss—is the most common form of liquidity in the literature. Funding liquidity, which is the ability to access funds to repay an existing loan, enters the literature as a complement to market liquidity in studies of the microstructure of dealers and market makers. That market and funding liquidity are liable to a fallacy of composition is recognised by categorising them as privately provided outside liquidity, in that the liquidity provided is not intrinsic to the asset in question.

Market and funding liquidity refer to the conversion of assets into monetary or pure liquid form. The former captures the idea of saleability or marketability of an asset, the latter the ability to acquire liquidity to postpone payments of existing debts. Each aligns with the necessity that, when a payment is due, offsetting credits must be available with which to make the payment. If there are insufficient credits then the agent must acquire them and may need to appeal to market or funding liquidity.

Two other options are available: either to rely on credits arriving into one's possession; or have on hand a ready store of credits that are due at that moment, that is: money. Thus, market making provides transactional liquidity associated with portfolio flows, and can be distinguished from liquidity arising from income-producing activities and liquidity flows from short-dated or self-liquidating assets. Inside liquidity, in the form of cash balances, is provided by assets with a price-protection feature that provide an overarching stability in the unit of account. Otherwise-safe assets such as government bonds are convenient in that they are readily shiftable but cannot provide capital certainty.

Table 2 – Forms of Liquidity Supply

Liquidity	**Inside/Outside**	**Private/Public**	**Supplier**	**Example**
Purely Public	Inside & Price-Protected	Public	Government	Central bank reserves and state money
Private-Public	Inside & Price-Protected	Hybrid	Banks	Insured bank deposits
Public-Private	Inside & Price-Protected	Hybrid	Money market dealers	Government bond-backed repos
Purely Private	Inside & Price-Protected	Private	Money market funds and banks	Uninsured deposits
Self-Liquidating	Outside	Either	Production	Bills of exchange
Trade	Outside	Private	Consumption	Sales of goods and services
Funding	Outside	Private	Banks	Borrowing
Market	Outside	Private[29]	Dealers	Sales of assets

What appears is a hierarchy of liquidity from price-protected money-claims, through trade liquidity, self-liquidating assets, funding liquidity, to market liquidity. The spectrum falls between the inside and the outside categories of pure cash and the reliance on the sale of assets. Other categories that must be considered are the public and private provision, and the availability in the face of aggregate and idiosyncratic disruptions. Table 2 provides a summary of the liquidity forms and spectrum thus identified.

[29] Except possibly for those assets that are part of the central bank's open-market operations.

The purpose of this and the previous two chapters was to conduct an ontological reflection on the nature of liquidity, money and the banking system. The characteristics of each were presented with as little theoretical interpretation as possible. In the case of money, however, it was not possible to avoid theory and the assumptions underlying each theory were identified and those theories that most aligned with historical and empirical evidence were preferred.

Part II has performed an ontological reflection on the characteristics of liquidity, money and the banking system. The resulting ontological landscape is used in the following chapters to assess existing theories for correspondence with reality. Part III discusses theories of liquidity in the tradition of Real Analysis, in which the neutrality of money is considered axiomatic. Then Part IV considers liquidity in theories in the alternative tradition of Monetary Analysis which retains a fundamental role for money. In Part V, elements of both Real and Monetary Analysis are combined to construct a coherent and holistic structure around liquidity.

Part III Real Analysis

It has long been recognised that incorporating a monetary asset into general equilibrium models is problematic. 'The most serious challenge that the existence of money poses to the theorist is this: the best-developed model of the economy cannot find room for it' (Hahn, 1982, p. 1). The difficulty arises from their basis in the tradition of 'Real Analysis', which

> proceeds from the principle that all the essential phenomena of economic life are capable of being described in terms of goods and services, of decisions about them, and of relations between them. Money enters the picture only in the modest role of a technical device that has been adopted in order to facilitate transactions. (Schumpeter, 1954, p. 277)

When analysis is conducted in real terms, and the neutrality of money is axiomatic, money becomes an afterthought. 'Monetary Analysis', by contrast, 'introduces the element of money on the very ground floor of our analytic structure and abandons the idea that all essential features of economic life can be represented by a barter-economy model' (Schumpeter, 1954, p. 278).

The exclusion of money from general equilibrium models is an oversight that inhibits the applicability of these models, especially with regard to the analysis of liquidity. More far-reaching consequences cannot be discounted since:

> It so happens that so far macro-economic theory has not come to serious grips with the phenomenon of liquidity preference in its widest sense. In particular, it has not allowed for transaction costs and uncertainties that arise for assets, especially real assets, other than money. That this is so is well illustrated by the popularity of the "representative" agent. This device abstracts from transactions that are the result of the heterogeneity of agents. Many propositions of macro-economics are seriously at risk from this neglect. (Hahn, 1990, p. 79)

The inability of general equilibrium models to incorporate money, and hence liquidity, in a fundamental sense is a serious deficiency. In Part III, the relationship between liquidity and money in the tradition of Real Analysis is considered. Chapter 6 outlines the two main strands of general equilibrium: the Wicksellian theory of the natural rate of interest and neo-Walrasian general equilibrium, as well as Irving Fisher's theory of the rate of interest.

Chapter 6 also considers the Rational Expectations Equilibrium and highlights the Wicksellian, Fisherian and Walrasian foundations beneath modern asset-pricing theory to identify its potential strengths and weaknesses. Chapter 7 evaluates the attempts to introduce liquidity and money into general equilibrium models.

6 General Equilibrium Theory

6.1 Introduction

General equilibrium theory can be divided into two theoretical strands: Wicksellian and neo-Walrasian. In assessing the usefulness of general equilibrium theory in explaining liquidity, we must first consider 'the distinction in theoretical structure between neo-Walrasian and Wicksellian theory. Both are neoclassical general equilibrium theories but both have distinct properties and hence distinct monetary theories' (Rogers, 1989, p. 5).[30] Neo-Walrasian theory operates in a world of perfect barter where all commodities are traded directly for each other. The Wicksellian approach, on the other hand, recognises a distinction between credit and money and introduces the concept of a 'natural rate of interest' as the keystone for its monetary theory. As a complement, Fisher's theory of interest can apply to either form of general equilibrium, and so its implications for liquidity, money and asset pricing must also be identified. As we shall see, each approach has fundamental difficulties incorporating money, and hence liquidity, into its framework.

This chapter proceeds in five main parts. The first reviews the Wicksellian theory of the natural rate of interest and its implied scarcity of liquidity. The second introduces Fisher's theory of interest based on time preference that provides the crucial element missing from the Wicksellian theory. The third part evaluates the relationship between liquidity and neo-Walrasian general equilibrium theory with its assumption of perfect market liquidity, or liquidity abundance. The fourth part considers the Rational Expectations Equilibrium and the fifth introduces the concepts underlying modern asset-pricing theory.

Although each form of general equilibrium stems from distinct theoretical foundations, the line of delineation between Wicksellian and neo-Walrasian models has become blurred in the literature, with Fisher's theory incorporated into either. Since each approach has its own difficulties explaining liquidity, the distinction is maintained during the discussion of asset pricing.

[30] A neoclassical or competitive equilibrium is one in which all optimal plans are achieved (Arrow & Hahn, 1971, p. 107).

6.2 Wicksell and the Natural Rate of Interest

Wicksell ([1898] 1936) analyses the distinction between bank credit (inside money) and money-proper (outside money). After considering the limits of pure credit creation, Wicksell subsequently postulates the existence of a unique 'natural rate of interest on capital' ([1898] 1936, p. 102), equivalent to the rate determined in a non-monetary economy, that ensures price stability. This rate of interest is derived from the productivity of capital, and 'would be determined by supply and demand if no use were made of money and all lending were effected in the form of real capital goods' (Wicksell, [1898] 1936, p. 102). As such, is it is based on Böhm-Bawerk's time-based theory of interest as a reward for the 'technical superiority of roundabout methods of production' (Conard, 1959, p. 39), where the real rate of interest is uniform across all assets and equal to the marginal productivity of capital.

Wicksell's innovation is to recognise the distinction between a market-determined rate of interest, which clears the credit or money market, and an underlying natural rate of interest determined solely by the real economy (Schumpeter, 1954, p. 1118). Wicksell reasons that, if banks were to offer credit money at a market rate lower than this natural rate, then 'prices [would] rise continually higher and higher' in a 'cumulative process' (Wicksell, [1898] 1936, p. 95) and, conversely, an indefinite fall in prices would result if the market rate were set above the natural rate. In an economy based on pure credit, this cumulative process would lead to price indeterminacy, implying that a commodity money is essential to place a binding reserve constraint on banks to prevent market rates from deviating from the natural rate (Patinkin, 1965, p. 368). Wicksell's conclusion is that, since prices are determinate in reality, the QTM is valid and a purely credit-based economy is impossible, in accordance with practical metallism. The supply of credit money is then restricted by the supply of the (scarce) outside or commodity money.

The theoretical and conceptual foundations for Wicksellian models are problematic. The first problem stems from the derivation of the marginal productivity of capital due to the difficulties in selecting a unit for capital inputs entering an aggregate production function. The existence of a unique Wicksellian natural rate of interest is not guaranteed unless capital and output are measured in terms of the same commodity or in units of monetary value (Rogers, 1989). The former cannot extend to a multi-commodity world and allows for technological liquidity to obscure other forms of liquidity; the latter involves treating

the rate of interest as both exogenous and endogenous simultaneously (Rogers, 1989, p. 38). Consequently, the Wicksellian system is a 'one-sided productivity model' (Hirshleifer, 1967, p. 196) and is incomplete without the addition of time preference. The productivity of the system can potentially be consistent with any externally determined interest rate (Hirshleifer, 1967).

To be complete, Wicksellian general equilibrium theory must be extended so that the natural rate of interest is determined by both the marginal productivity of capital and intertemporal consumption preferences. In this form of the 'loanable funds' theory, the natural rate of interest serves to equilibrate the flows of real investment and real saving (Leijonhufvud, 1981), and is unaffected by monetary policy, thereby placing Wicksellian theory in the tradition of Real Analysis. Wicksellian theory attempts to reconcile the liquidity abundance associated with pure credit with the liquidity scarcity view of the QTM by means of an insuperable, but unobservable, equilibrium rate of interest determined independently of the monetary system. Consequently, 'the natural rate of interest lies at the heart of a Wicksellian statement of the loanable funds theory and both stand or fall together' (Rogers, 1989, p. 22).

In Wicksellian theory, 'disequilibrium in the goods market, usually expressed as a difference between planned saving and investment, affects the rate of interest' (Hayes, 2010, p 808; see also Robertson, 1940, p. 18). It predicts that 'a rise in thrift *directly and immediately* lowers the rate of interest unless obstructed by either one of two specific factors, namely, a simultaneous rise in hoarding or by credit contraction' (Bibow, 2001, p. 593, original emphasis). In the loanable funds view, an increase in saving means that savers either buy bonds—that is, supply loanable funds—so that interest rates fall, or they hoard deposits so that the interest rate stays unchanged (Bibow, 2001, p. 597). Difficulties with this prediction appear when the implications for a rise in thrift, which has an impact on buffers of goods for sale, are applied to the unavoidable financial counterpart (Bibow, 2001, p. 595). The loanable funds theory overlooks the equal and opposite demand for funds by the firms that need to finance their inventories (Bibow, 2001, p. 598). For firms, the result of an unexpected drop in sales is a build-up of unsold inventories, or unplanned investment, and the consequent requirement for unexpected funding. Although interest rates may well change to meet revised portfolio preferences, this change is only an indirect result. Depending on liquidity preference (see Chapter 8), interest rates may well rise (Bibow,

2001, p. 599). The inaccuracy of this prediction signifies an 'analytical flaw' such that 'loanable funds theory is *logically inconsistent* and should thus be abandoned' (Bibow, 2001, p. 592, original emphasis).

Furthermore, 'real' saving in the Wicksellian sense is undefined in a world where capital goods are produced (Rogers, 1989, p. 42), especially when it is recognised that savings are intermediated in the modern economy (Chick, 1983, pp. 56-57). Banks do not simply intermediate prior savings; they create money. Banks provide finance for investment via money creation, and do not just intermediate pre-existing savings. As seen in Chapter 3, the hoarding of money cannot result in funds that are unavailable for investment. The hoarded money is already invested, due to the very nature of its existence (or the very fact that it exists).

Loanable funds theory ignores the ability of banks to create liquidity and money '*instantaneously* and *discontinuously*' (Jakab & Kumhof, 2015, p. 4, original emphasis) by imposing the restriction that they can only 'lend' pre-existing savings. In reality, 'cash is never lent directly but only withdrawn against deposits that have first been created through lending' (Jakab & Kumhof, 2015, p. 5). The Wicksellian theory, being in the tradition of Real Analysis, has no fundamental place for either liquidity or money. 'It turns out that loanable funds proponents simply confuse *money* and *saving*' (Bibow, 2001, p. 604, original emphasis). *Ex-ante* investors are assumed to finance their investment out of *ex-ante* savings, a concept that has no place in a monetary economy.[31] Savings can only 'finance' investment in a single-commodity non-monetary world.

For Bibow (2001, p. 610, original emphasis) 'the Wicksellian notion of *the* "natural rate of interest", as determined by the real forces of productivity and thrift, is irreparably undermined as the anchor of the system.' Rather than the market rate of interest adjusting to restore equilibrium with the 'natural' rate, Rogers (1989, p. 212) proposes that the interpretation should be reversed. An increase in the market rate is followed directly by a 'fall in the spot prices of all durable assets' that equilibrates expected returns with the new market rate (Rogers, 1989, p. 213). Consequently, changes in the market rate of interest

[31] That it is not possible for *ex-ante* saving to finance *ex-ante* investment undermines both the Hicks/Ohlin version of loanable funds—where planned savings are from future income—and the Robertson version—where planned savings are from past income (Robertson, 1940, p. 6). For discussion, see Hayes (2010).

affect the flow of investment, the level of speculative loans and possibly contribute to asset-price bubbles (Turner, 2013). Wicksell's focus on the effect of credit creation on price stability should be broadened. It follows then that the rejection of the natural rate of interest requires more care when interpreting the direction of causation of Fisher Equation, which ultimately is a definition of the real rate of interest. It is Fisher's theory of interest that we must consider next.

6.3 Fisher's Theory of Interest

The addition of time preference to Wicksellian theory requires a clarification of the relationship between the natural rate of interest and Fisher's theory of interest (Fisher, 1896). Fisher's theory of the rate of interest is a market-determined 'real' rate of interest that aligns consumer time preferences with technological production possibilities, under conditions of full employment (Cottrell, 1994b, p. 428) and constant prices (Conard, 1959, p. 140). The interest rate so determined is 'real' in the sense that it represents the trade-off between consumption of a single commodity in the present and in the future. Since this trade-off is conducted without money, the theory is also in the tradition of Real Analysis. Furthermore, because the rate of interest established is simply a ratio of commodity prices, unrelated to capital or time, it is doubtful that it is a theory of interest at all (Rogers, 2007, p. 142). The rate of interest is not guaranteed to be positive (Rogers, 2007, p. 140).

Confusingly, however, Fisher defines the real rate of interest r as the nominal rate of interest i adjusted for inflation[32] π, giving, in approximation, the Fisher Equation:

$$i \approx r + \pi. \tag{6.1}$$

The real rate of interest from equation (6.1), being the nominal rate of interest adjusted for the change in the value of money, is conceptually different from the 'real' rate of interest derived from either the Wicksellian or Walrasian forms, or even Fisher's theory of the rate of interest (Keynes, 1930a, p. 197, n. 2; Rogers, 1989, pp. 21-22, n. 1; Smithin, 2006).

The definition of the real rate of interest and the axiom of money neutrality leads to the 'Fisher Effect' whereby, in equilibrium, the nominal interest rate adjusts one-for-one with

[32] The real rate of interest can be *ex ante* or *ex post* depending on whether, respectively, expected or realised inflation is used in the equation (Conard, 1959, p. 140; Fisher, 1896)

changes in expected inflation. There are several problems with this effect. First, in many cases, anticipated future price rises will be reflected immediately in current prices, before any effect on the money rate of interest can take place (Keynes, 1936, p. 142). A subtler problem, but of relevance to this study, is that changes in expected inflation will result in the same loss of purchasing value on non-interest paying money balances, as on interest-bearing assets. The rate of interest is not a return for holding money, it is a return for lending it: the expectation of a change in the value of money will fall equally on money and loans of money (Cottrell, 1994b, p. 427; Kaldor, 1982, p. 97). At issue is the relative desirability of holding money, since its real value will be reduced by the expected rate of inflation $-\pi^e$ versus lending money, which earns $i - \pi^e$ in real terms. Inflation has an equal impact on the two alternatives and can therefore have no influence on the decision to hold or lend money. That the nominal rate adjusts to anticipated rates of inflation 'is false for the simple reason that the holding of liquid financial assets (which is the alternative to holding bonds) is exposed to exactly the same risk of erosion in real value through inflation as gilts [UK government bonds] are' (Kaldor, 1982, p. 97).

A deeper issue concerns the increase in investment that may result from an increase in inflation expectations (Cottrell, 1994b, p. 420). With interest rates unchanged, higher inflation expectations will increase the discounted value of future profits and, if production prices of investment goods are unchanged, increase investment itself. Is there a higher nominal rate of interest that would offset this increased investment, and what is its relationship with the increase in inflation expectations? The Fisher Effect states that the new nominal rate will be derived from the previous real rate. The problem is, inflation affects money and money lending in the same way, therefore only a higher real rate can offset the increase in investment. The Fisher Effect is incorrect: money rates cannot adjust to inflation 'directly', a 'mistake' identified by Keynes (1936, p. 142). A higher real rate of interest is needed to offset the increase in investment, but the Fisher Effect does not ensure this outcome. The required increase in the real interest rate cannot be the result of a greater demand for loanable funds, since the real rate of interest is unchanged by assumption in Fisher's theory.

Neo-Wicksellian models overlook these difficulties and use the Fisher Equation as a means to derive expected inflation from the Wicksellian natural rate of interest and the central bank's nominal policy rate. Wicksellian theory is widely used in the neoclassical monetary

literature as the basis of a theory of interest rates (Galí, 2008; Woodford, 2003). By the premise of the QTM, monetary policy conducted by interest rates, and not the stock of money, raises questions of price determinacy (Black, 1970b; Sargent & Wallace, 1975). Combining the Fisher Equation with the Wicksellian natural rate of interest provides the foundation for the Taylor rule (Taylor, 1993) determining how central banks should conduct interest-rate policy. A central bank following the Taylor rule will automatically adjust nominal interest rates in response to any deviation of inflation from its target rate, thereby forcing market interest rates to diverge from the natural rate of interest. In this way, neo-Wicksellian models, where Wicksellian ideas are coupled with Rational Expectations, allow for the complete absence of either outside or inside money while still preserving price determinacy in a cashless model (Woodford, 2003; see Cochrane, 2011a, for a critique.). Similarly, the neo-Fisherian literature places the Fisher Equation onto a Wicksellian foundation to reverse the commonly accepted 'stylised fact' that raising policy rates lowers inflation (Cochrane, 2017b). The fixed natural rate of interest implied by the Wicksellian monetary foundations contradict this stylised fact, since the product of higher nominal interest rates can only be 'an immediate and permanent rise in expected inflation' (Cochrane, 2017b, p. 3).

Nevertheless, loanable funds theory and the Fisher Equation underpin the central banks' current consensus regarding the use of a negative interest rate policy (NIRP) as a means to increase aggregate demand and to combat the a 'global savings glut' (Bernanke, 2005) or 'secular stagnation' (Summers, 2016). In either case, it is estimated that the Wicksellian natural rate[33] is negative. Pre-NIRP, the consensus was that central banks were restricted by a zero lower bound (ZLB) on interest rates. The contents of the pre-NIRP toolkit were: QE; 'forward guidance', promising lower policy rates in the future (Woodford, 2012); or to 'credibly promise to be irresponsible' so that inflation expectations increase thereby lowering real interest rates via the Fisher Equation (Krugman, 1998). The introduction of negative interest rates and the breach of the ZLB have prompted a revision of this consensus. Unfortunately, the NIRP has had very limited apparent success in stimulating aggregate demand, and its Wicksellian theoretical basis renders it 'profoundly wrong'

[33] The Wicksellian natural rate of interest is often represented by r^*. For example, see Laubach and Williams (2003).

(Palley, 2016, p. 2).

6.4 Walras and Pure Exchange

Neo-Walrasian[34] general equilibrium theory (Arrow & Hahn, 1971; Debreu, 1959) identifies the assumptions required in a hypothetical economy for all markets to clear simultaneously. One such assumption is a complete set of Arrow-Debreu contracts available for all commodities, trading dates and future states of nature, also known as complete markets. These contracts enable agents to conduct all conceivable transactions at the beginning of time, time-0, thereby reducing the economy to frictionless pure exchange. Briefly, the process of arriving at equilibrium is achieved by a fictional Walrasian auctioneer who, in an 'as if' procedure called *tâtonnement*, facilitates multilateral exchanges between all agents until a Pareto efficient allocation[35] is achieved. Agents are allowed to recontract during this process, but, by assumption, the final set of exchanges is honoured without default. Much of the literature concentrates on establishing the conditions for the existence, uniqueness and stability of this equilibrium.

The result, since any commodity can be exchanged for any other commodity (including in any future time or location), is that 'under an Arrow-Debreu auction all commodities are equally liquid' (Rogers, 2008b, p. 19). In terms of the taxonomy of liquidity outlined in Chapter 5, all commodities display perfect market liquidity. A consequence of perfect market liquidity is that there is no need for any form of physical money beyond the *numéraire*, a commodity arbitrarily chosen to represent the unit of account. Money loses its form as either a medium of exchange or a store of value. In the absence of uncertainty about future commodity flows, there is no requirement for liquidity or a liquid asset. Perfect market liquidity renders all other forms of liquidity irrelevant; they are disregarded and entirely absent. It can be argued that liquidity is only a property of a money-mediated system; a pure-exchange system has no concept of liquidity (Lozano, 2015, p. 22). With complete markets, all agents can risk share to an extent that justifies the use of a single

[34] A Walrasian general equilibrium is based on Walras' Law, which states that the sum of the excess demands and excess supplies of all goods must be zero (Niehans, 1978, p. 9-12; Patinkin, 1965, pp. 35-36).

[35] A Pareto efficient allocation is a set of feasible commodity allocations between agents for which 'there is no way of making everyone better off' (Arrow & Hahn, 1971, p. 91).

‘representative’ agent, a modelling technique which itself suggests an absence of financial transactions.

The emphasis given in Chapter 3 to the distinction between the unit of account and monetary objects is reflected in the concept of the *numéraire* in Walrasian theory. A key difference between the *numéraire* and an abstract unit of account, however, is that the *numéraire* is a commodity (or fiat money treated as an ‘as if’ commodity), which has an arbitrary price of unity. An Arrow-Debreu contract is a promise to deliver real goods, not money—it is not a financial contract. As such, the *numéraire* is tied to the tradition of Real Analysis and expresses the commodity view of money. The unit of account, on the other hand, is an abstract measure with an uncertain real value and no corresponding physical presence used to denominate financial assets.

The axiom of money neutrality allows for the analysis of the economy within the context of pure exchange without liquidity or money. The net-money doctrine (see Section 3.2) is used to justify the simplification of analysing models without an explicit banking system or even money itself (Woodford, 2003) and the absence of an integrated theory of liquidity. The Modigliani-Miller theorem (see Section 5.7) is an expression of the net-money doctrine and leads to the conclusion that bank intermediation and inside money cannot have significant macroeconomic effects. By applying the Modigliani-Miller theorem to the banking system, Fama (1980, p. 44) denies that bank deposits are money, and can neither be the *numéraire* nor affect the problem of price level determinacy. The result is that ‘as portfolio managers, banks are financial intermediaries with no special control over the details of a general equilibrium’ (Fama, 1980, p. 45). Only outside currency issued by the government or central bank reserves that reduce transactions costs can perform this function (Fama, 1980, pp. 50-53).

With pure exchange, banks—including central banks—lose any differentiating factor that could explain their macroeconomic effects or even their very existence. Wallace (1981) highlights the confusion that the net-money doctrine causes by applying the Modigliani-Miller theorem to open-market operations to demonstrate the irrelevance of government portfolio constructions and ends up questioning why government bonds cannot be ‘spent’. Models where central bank profits, or *seigniorage*, are included in a representative agent’s budget constraints struggle to show that open-market operations have impact due to a lack of any wealth effect (Barro, 1974). Similarly, no explanation can be provided for ‘one of

the most obvious “liquidity” premiums: Money is overpriced—it has a lower discount rate—relative to government debt, though they are claims to the same payoff in a frictionless market’ (Cochrane, 2011b, p. 1079). By analysing QE in such a liquidity-free framework, Ben Bernanke concludes that ‘the problem with QE is it works in practice, but it doesn’t work in theory’ (Ahamed & Bernanke, 2014, p. 12).

Buiter (2009) doubts that the assumption of complete markets is an ‘appropriate simplification’, since it implies an infinite number of markets that would ‘exhaust the resources of the universe’. For Lucas (1984, p. 13), however, the inconsistency with reality of time-0 trading and complete market of securities is ‘superficial’. Instead, Rational Expectations and the no-arbitrage condition ensure that the prices arising from more frequent trading agree with the time-0 prices, and reduce the need for superfluous contracts (Lucas, 1984, p. 14). The requirement for complete markets and clearing at time-0 can be replaced by sequential complete markets (Harrison & Kreps, 1979; Ljungqvist & Sargent, 2004, pp. 223-230).

Finally, it must be noted that the neo-Walrasian system produces the prices of commodities dated by delivery time. Interest rates can only be extracted as the ratio of the prices of like commodities. As such, these rates are real, but not in the sense that they are nominal rates adjusted for purchasing power using the Fisher Equation. There are as many real rates of interest as there are commodities and delivery times; the neo-Walrasian framework provides a theory of value, not a theory of the rate of interest. Interest rates derived from the prices of time-dated commodities provide no ‘analytical insights’ beyond those provided by the prices themselves (Rogers, 1989, p. 52). As in the Fisherian theory of interest, time is no different from location in a time-0 auction and there is no guarantee that neo-Walrasian real interest rates are positive. Nor does a neo-Walrasian model satisfy the Wicksellian condition that the rate of interest is equal for all assets in equilibrium.

The distinction between the Wicksellian and neo-Walrasian theories can be seen from their treatments of liquidity. Mehrling (2000b, p. 15) pinpoints the issue as a ‘fundamental source of misunderstanding’ which manifests itself in the ‘difference between the liquidity scarcity view of economics and the liquidity abundance view of finance.’ Liquidity scarcity is expressed by the Wicksellian notion that market interest rates must adjust to agree with the natural rate of interest defined in the real economy. Thus,

> …the persistent attraction to economists of the quantity theory of money, even

> given all its faults, is that it expresses succinctly the economist's intuitive sense that the real liquidity of the economy as a whole is scarce and that attempts to increase liquidity by expanding nominal money must eventually reckon with this fundamental real scarcity. (Mehrling 2000b, p. 15)

On the other hand, the neo-Walrasian 'theory of value abstracts from the scarcity of liquidity and treats all commodities as equally and perfectly liquid' (Mehrling 2000b, p. 15). These views are incompatible, both with each other and with the ontology of either liquidity or money. Consequently,

> …monetary theory cannot be brought under the theory of value (as currently constructed) but no alternative analytical structure has yet emerged to gain general acceptance. Until it does, the old distinction between money and credit remains as the economist's crude theoretical attempt to grapple with the apparent hierarchy of liquidity, with credit viewed as a mechanism for stretching scarce liquidity (money). Similarly, the old attempt to measure the quantity of money remains as the economist's crude empirical attempt to grapple with the same hierarchy by measuring the size of its base. (Mehrling 2000b, p. 15)

Both the 'economics view' and the 'finance view' (Mehrling, 2013b, p. 356) in their attempts to understand the world 'resolutely abstract from money' (Mehrling, 2013b, p. 357) in the perfect exchange constructions of neo-Walrasian general equilibrium ('trade in goods') and the perfect liquidity modern asset pricing ('trade in financial assets'). Ignoring the monetary or liquidity dimension and 'the standard practice of focusing attention on the ''equilibrium'' or ''fundamental'' exchange rate amounts to focusing on a special limiting case that would prevail if matched-book and speculative dealers were willing to do their work for free' (Mehrling, 2013b, 358).

The acknowledgement of this omission has resulted in various attempts to incorporate money into general equilibrium theory. These attempts are not entirely successful in providing explanations, as we shall see in the next chapter. Before that, we must look more deeply at Rational Expectations and modern asset pricing, with their failure to explain the observed frequency of trading and assumptions of perfect market liquidity.

6.5 Rational Expectations

The Rational Expectations Hypothesis (REH) stems from the straightforward idea that

‘expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory’ (Muth, 1961, p. 316). The precise statement of the hypothesis distinguishes between subjective and objective probabilities, since the

> …expectations of firms (or, more generally, the subjective probability distribution of outcomes) tend to be distributed, for the same information set, about the prediction of the theory (or the “objective” probability distributions of outcomes). (Muth, 1961, p. 316)

The REH is qualified with two caveats. First, the ‘expectations of a single firm may still be subject to greater error than the theory’, and, second, the hypothesis does not ‘state that predictions of entrepreneurs are perfect or that their expectations are all the same’ (Muth, 1961, p. 317). The REH has evolved to be ‘the imposed extra assumption that the subjective probability distribution of outcomes believed by agents within an economic system equals the objective frequency distribution actually generated by the system itself’ (Weitzman, 2007, pp. 1102-1103). Conveniently, agents with Rational Expectations ensure the existence, uniqueness and stability of a Rational Expectations Equilibrium (REE) equivalent to a fully informed neo-Walrasian general equilibrium (Grossman, 1981). In a REE, the requirement for complete markets is removed but at the cost of ‘unlimited computational capacity’, without altering the conclusion that the theory has ‘no role for money and liquidity’ (Radner, 1968, p. 31). If uncertainty is a necessary requirement for a liquid store of value, then the REH removes the need for ‘the constant revision of strategies’ and continuous trading and, hence, the ‘demand for money and liquidity’ (Radner, 1968, p. 36).

The importance of the REH is in its implications for models in which agents do not have the same information set because access to information is asymmetric. For example, in models with asymmetric information, each agent may have a different estimate of each asset’s payoff. By means of the REH, private information available to only some agents can become public by means of its observed effect on asset prices. In a REE, uninformed investors can learn from the informed investors by conditioning their asset allocations on asset prices. The result is that ‘each agent has a different (but correct) asset pricing model’ conditional on their own information set (Admati, 1985, p. 640). Each agent’s expectations of future prices will differ, but this ‘creates an incentive for the opening of a futures

market...where they can bet against each other' (Grossman, 1981, p. 546).

When information is asymmetric, a REE improves on a naive neo-Walrasian equilibrium by allowing agents to use the private information of other agents so revealed in market prices, such that a REE is equivalent to a fully informed neo-Walrasian equilibrium. To establish a REE, 'traders need only know the stochastic process generating the equilibrium price' (Grossman, 1981, p. 545). Differences in information between agents are eliminated, since 'agents make statistically correct inferences on all the information they possess, including current prices' (Admati, 1985, p. 629), whereas neo-Walrasian equilibrium prices would prompt re-contracting by agents as soon as they were observed (Grossman, 1981, p. 549). The resulting 'equilibrium price is a random variable', not a number like in a neo-Walrasian equilibrium (Grossman, 1981, p. 544), in that the REE price vector clears the market for every future state of nature. This perfect foresight regarding the price in each state of nature is equivalent to a complete set of Arrow-Debreu contracts (Cass & Shell, 1983, p. 201). The result is that equilibrium prices are the same as if all investors were fully informed and information symmetry is restored.

Observed market prices can incorporate information to eliminate trading advantages, 'but if traders have diverse information sets, then these expectations need not be the same across traders. Thus, as in microstructure models, the adjustment of prices to full information values can differ widely across markets that are deemed efficient' (O'Hara, 2003, p. 1351). Price discovery, which is 'the incorporation of new information into asset prices' (O'Hara, 2003, p. 1339), must be distinguished from the cost of market liquidity or transactions costs. 'The symmetric information-based asset pricing models do not work because they assume that the underlying problems of liquidity and price discovery have been completely solved' (O'Hara, 2003, p. 1335). When information is asymmetric, or only partially revealed, in an otherwise neo-Walrasian model, then idiosyncratic risk requires a risk premium. The risks associated with price discovery and asymmetric information can affect expected returns but is an issue that is only peripherally related to liquidity.

The use of price information differentiates a REE from a neo-Walrasian equilibrium and formalises the idea of agents forming opinions of other agents' opinions as in Keynes's beauty contest analogy (Keynes, 1936, p. 156). The cost of gathering this information, however, makes the existence of a REE uncertain. Gathering information requires resources, which removes the incentive for each investor to gather information, thereby

undermining the foundations for efficient markets and the very existence of a REE (Grossman & Stiglitz, 1980). Of particular relevance to liquidity, the conceptual problems with REE are only solved by the addition of some form of 'noise' trading (Admati, 1985; Black, 1970b). Noise trading takes the form of a random supply of assets, which represents an unexplained random demand for assets by uninformed or liquidity-motivated investors (Admati, 1985, 632). 'Understanding liquidity requires us to unravel the puzzle of why real people and institutions trade so much more than they do in our models' (Cochrane, 2011b, p. 1071).

The relationship between the REH and the Efficient Markets Hypothesis (EMH) can now be identified. The EMH states that 'a market in which prices "fully reflect" available information is called "efficient"' (Fama, 1970, p 383). Furthermore, Fama (1991, p. 1575) divides the EMH into the 'extreme version of market efficiency', which is 'the simple statement that security prices fully reflect all available information', but 'is surely false' because of the costs of obtaining so much information. Instead a 'weaker and economically more sensible version of the efficiency hypothesis says that prices reflect information to the point where the marginal benefits of acting on information (the profits to be made) do not exceed the marginal costs'. In other words, opportunities for profit based on publicly available information, such as past asset-price behaviour and dividend forecasts, will be eliminated where possible. The implication of the EMH is that technical trading (for example, the use of charts of past prices to predict future prices) is ultimately futile. This translates into the idea that it is impossible to beat the market, the market is unpredictable, but not necessarily that the market is always 'right'.

Jespersen (2009, p. 69) highlights the metaphor associated with the REH and proposes 'ideal' expectations as a better description to expose its opposition to 'realistic' expectations. A REE is a much stronger proposition than an equilibrium where all agents' expectations are in alignment and future prices are 'agreed' but need not be correct, just consistent across all agents (Hayek, 1937). Therefore, the

> REE is a seriously misleading equilibrium concept for pricing assets because it is describing an unstable knife-edge balance in price distributions, having probability-of-existence measure zero, which unravels completely in the presence of even an infinitesimally small bit of evolutionary-structural uncertainty. (Weitzman, 2007, p. 1115)

It is important to question the REH, without which the concept of fundamental asset prices established by a general equilibrium is impossible. The full risk sharing implied by general equilibrium is unrealistic: 'Risk sharing needs *everyone* to change their portfolios and bear a risk in order to eliminate segmentation' (Cochrane, 2011b, p. 1072, original emphasis). Geanakoplos (1990, p. 16) clarifies that efficient markets in incomplete markets models do not imply 'efficient or Pareto optimum allocations', merely the existence of a pricing kernel such that all the prices are subject to the law of one price (referred to as 'no arbitrage'). The pricing kernel 'typically does not correspond to any agent's measure of beliefs, even when they all agree' (Geanakoplos, 1990, p. 16).

The risk that a price deviates from its fundamental value is eliminated by complete markets or the REH. When all assets have perfect market liquidity, interest rates and asset prices are determined by intertemporal preferences. Even in explicitly incomplete market models, agents are often still assumed to have 'perfect (conditional) expectations', that is, Rational Expectations, which ensure a general equilibrium rather than a temporary equilibrium (Geanakoplos, 1990, p. 10). Other simplifications, such as the existence of only one consumption good, eliminate the possibility of short selling and reduce the complexity of an incomplete market model to a complete market model (Geanakoplos, 1990, p. 11).

The difficulty that incomplete market models have with short selling and borrowing is troublesome given the ontological nature of money as both a credit and debit on agents' balance sheets. As noted in Section 6.4, under the assumption of perfect market liquidity, markets can be made complete by continuous trading of long-lived assets, instead of Arrow-Debreu time-0 market clearing. Neo-Walrasian general equilibrium theory permits the analysis of an economy where 'all assets have well-defined prices' but where 'transactions are costly' (Hahn, 1990, p. 65). General equilibrium, by providing a solid platform of fundamental prices, restricts the focus to market liquidity, thereby obscuring all other forms of liquidity.

For example, Brunnermeier and Pedersen (2009) use the difference between the market price and fundamental value to represent market illiquidity (p. 2210). This difference is the source of profit for the 'traders' that provide market liquidity. Information asymmetry means that only some agents are aware of the fundamental value, and, because this expected profit is not known to all agents, 'illiquidity' exists in the model. Traders are partially debt-funded by financiers, using assets as collateral. The entire value of the asset cannot be

borrowed, so that traders must provide their own funds to cover a margin or 'haircut' that financiers vary with observed asset-price volatility. The margin increases with volatility because financiers do not know expected profits, and consequently funding illiquidity and market illiquidity move in the same direction. Conversely, and counter-intuitively, if financiers had complete information they would charge a margin that falls with higher volatility and 'illiquidity' (Brunnermeier & Pedersen, 2009, p. 2220).

Similarly, for Shleifer and Vishny (1997), noise traders generate a difference between an asset's market price and its fundamental value that represents potential arbitrage profits. Information asymmetry, however, means that financiers cannot observe the potential profits of arbitrageurs, only price volatility. The result is that arbitrage funding is removed when potential arbitrage profits are highest, causing involuntary liquidations and further moves from fundament value (Shleifer & Vishny, 1997, p. 46). Without a large number of diversified arbitrageurs, idiosyncratic risk affects the funding of strategies and undermines the 'theoretical underpinnings of the efficient markets approach to arbitrage' (Shleifer & Vishny, 1997, p. 52). Again, the model relies on the introduction of noise or sentiment traders who are unable to understand the strategies of the arbitrageurs (Shleifer & Vishny, 1997, p. 46).

As stated in Section 5.2, attempting to extract profits from the difference between market values and fundamental values is the activity of a speculator and not a market maker. In noise-trader models, by contrast, informed speculators work as combined market makers and outside-spread investors. Information asymmetry and market liquidity are the predominant features of attempts to add liquidity to general equilibrium models (O'Hara, 2003). In a general equilibrium model, market makers need to be wary of counterparties with private information and that have an advantage in assessing fundamental value (Amihud et al., 2005, pp. 295-298; Kyle, 1985). Models based on information asymmetry with reference to a fundamental or 'true' underlying price (see, for example, Ho & Stoll, 1981, p. 48) assume behaviour not reflective of actual dealer activity.

A reliance on fundamental value reinforces the need for *ad hoc* assumptions of exogenous price shocks, or noise traders, which randomly move prices from their fundamental value. Without noise traders, those who are misinformed, there would be no trading, 'noise trading is essential to the existence of liquid markets' (Black, 1970b, p. 529). Furthermore, 'the noise traders as a group will lose money by trading, while the information traders as a group

will make money' (Black, 1970b, p. 531). By aligning liquidity with market liquidity within a general equilibrium model, the whole purpose of trading disappears and an unexplained random element must be introduced.

In general equilibrium, where the system is converging on a fixed point, the long-period point of convergence is known, and the 'equilibrium itself is independent of the traverse' or starting point (Jespersen, 2009, p. 11). Initial conditions do not matter, only the equilibrium point itself is significant. Agents are surprised at short-term events, but the long-term is certain. To achieve greater realism 'the macroeconomic process can only be perceived within an open framework' (Jespersen, 2009, p. 50) where the end-point is unknown. With this level of uncertainty, then the long-period point does not exist and the system becomes 'open' and its path-dependence creates even more uncertainty. The system diverges from a fixed point and the traverse itself becomes the focus of attention. Agents behave rationally with limited information and inherent uncertainty. They have more certainty about the short term; the long term is much more uncertain.

Asset-pricing theory does provide, however, a useful framework within which to discuss liquidity, money in both Real and Monetary Analysis. The relevant aspects of modern asset-pricing theory are discussed in the next section.

6.6 Asset Pricing

Wicksellian, neo-Walrasian and Fisherian concepts merge in modern treatments of asset pricing in the finance and economics literature. The assumption of complete markets, whereby an Arrow-Debreu contingent-claim contract exists for every conceivable future state of nature, leads to perfect market liquidity, and, consequently, the exclusion of any other form of liquidity. The complete set of Arrow-Debreu contracts also ensures a neo-Walrasian general equilibrium and risk sharing between representative agents. The possibility of risk sharing also allows a risk-free asset to be created from a combination of contingent claims. This risk-free asset plays a pivotal role in asset pricing and removes the need for a liquid asset. In this section, both the generalised asset-pricing model and the specific consumption-based asset-pricing model are presented. The generalised model frames the discussion of liquidity and money in following chapter. The consumption-based model is used to consider liquidity in asset pricing.

The application of general equilibrium models to the problem of determining asset prices

and interest rates is based on an investor's expected utility trade-offs, via a 'stochastic discount factor'. In the case of complete markets, or the Rational Expectations equivalent, the stochastic discount factor is unique and investors are able to share risk to produce an optimal resource allocation (Cochrane, 2005b, pp. 54-56). The additional assumption of a single consumption good allows asset returns to reflect the marginal productivity of capital, and the Fisherian subjective time preference ties asset pricing to the intertemporal consumption choices of the representative agent. The agent, in assessing the decision to consume now or save for the future, arrives at a point where the marginal utility of present consumption is balanced by the marginal utility of future consumption.

In a simple derivation of the consumption-based model[36], a two-period investor can buy or sell as much of an asset with payoff x_{t+1} as they wish. In addition, the investor receives endowments e_t and e_{t+1} and enjoys consumption c_t and c_{t+1} according to time-separable utility function:

$$U(c_t, c_{t+1}) = u(c_t) + \mathbb{E}_t[\beta u(c_{t+1})]. \tag{6.2}$$

Where β is the subjective discount factor and $\mathbb{E}_t[.] = \mathbb{E}[.|\mathcal{F}_t]$ is the (subjective or objective) expectation conditional on $\mathcal{F}_t$, the set of information known by the investor at time t.

The optimal amount that the investor would purchase ξ solves the objective function:

$$\max_{\xi} \; u(c_t) + \mathbb{E}_t[\beta u(c_{t+1})] \tag{6.3}$$

subject to the constraints

$$c_t = e_t - p_t\xi, \tag{6.4}$$

and

$$c_{t+1} = e_{t+1} + x_{t+1}\xi. \tag{6.5}$$

Where p_t is the asset price in terms of the single consumption good at time t.

Conducting substitutions and differentiating with respect to ξ gives the first-order condition:

[36] This derivation is taken from Cochrane (2005b, pp. 5-6). For a more complete treatment, see Ljungqvist and Sargent (2004).

$$p_t = \mathbb{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right]. \tag{6.6}$$

The asset price p_t is established at the point where the expected marginal utility of consumption at time $t+1$ is equal to the marginal cost of utility forgone at time t.[37] Equation (6.6) is a form of the Euler equation, the necessary first-order condition for optimal consumption.

The stochastic discount factor or pricing kernel m_{t+1} is then defined as

$$m_{t+1} \equiv \beta \frac{u'(c_{t+1})}{u'(c_t)}. \tag{6.7}$$

In infinite-horizon models, both the Euler equation (6.6) and a transversality condition are necessary conditions for optimal consumption; the Euler equation alone is sufficient in finite-horizon models. The transversality condition is generally expressed (Cochrane, 2005b, p. 27) as

$$\lim_{j \to \infty} \mathbb{E}_t [m_{t+j} p_{t+j}] = 0. \tag{6.8}$$

The transversality condition is an extra first-order, no-Ponzi condition that prevents the representative agent from borrowing an infinite amount in the present and rolling over these debts forever. Conversely, it prevents the agent from 'overaccumulating assets' since 'a higher expected lifetime utility could be achieved by, for example, increasing consumption today' (Ljungqvist & Sargent, 2004, p. 394). Another interpretation of the transversality condition is that it ensures that all assets are held by someone (Ljungqvist & Sargent, 2004, pp. 401-402). A strictly negative value would create an infinite demand for the asset; alternately no agent would demand the asset if the transversality condition were strictly positive. The problem with the transversality condition is that it specifies convergence to a fixed point, thereby closing the system and making the future more certain than the present. The transversality condition ensures that 'the influence of the infinitely distant future on asset prices today vanishes' (Buiter, 2009).

---

37 Shackle (1955, p. 116) uses the labels 'possessor-satisfaction' and 'consumer-satisfaction' to describe this trade-off.

Some observations are warranted. First, the Fisherian pedigree of the stochastic discount factor is evident in its being the marginal rate of substitution between consumption across the two periods. With complete markets, all risk is shared and the consumption level entering the marginal utility is effectively aggregate, figuratively consumed by a representative agent. Second, the payoff, and hence the asset 'price', is in terms of the consumption good, acting as *numéraire,* and not in terms of money. Finally, with complete markets the stochastic discount factor is unique and establishes unique prices for all assets. If, however, markets are incomplete, then there are an infinite number of valid discount factors (Cochrane, 2005b, p. 66).

An important assumption underlying the theory is that the amount an investor can transact is limitless, that is, perfect liquidity. Thus, 'present theories of asset pricing pay no attention to the liquidity characteristics of assets' (Hahn, 1990, p. 65). Arbitrage is the condition that $m_{t+1} > 0$, which ensures that any positive payoff has a positive price. Perfect arbitrage assumes perfect liquidity (Mehrling, 2011, p. 90) and so the effects of liquidity and money on asset pricing are assumed away. As a real-world consequence, Mehrling (2011, p. 65) claims that, prior to the GFC, liquidity was essentially regarded as a free good, with the associated allocation distortions.

Although the 'consumption-based model is, in principle, a complete answer to all asset pricing questions', it 'works poorly in practice' (Cochrane, 2005b, p. 41). Tests against the marginal utility of consumption are less than convincing (Cochrane, 2005b, pp. 41-43). Its predictions result in the equity premium and risk-free rate puzzles, where the theoretical risk premium on equities is too small relative to the observed premium and the risk-free rate is too large (Mehra & Prescott, 1985; Weil, 1989). It also cannot explain the term premium or volatility of long-term bonds (Holmström & Tirole, 2001, p. 1839). This disconnection between empirical evidence and theory has left a divide between traditional finance, which continues to focus on the practical task of relative asset pricing, and economics grappling with a theoretical explanation of absolute pricing using macroeconomic factors (Cochrane, 2005b, p. xiv). As such, opinions on the applicability of the Euler equation to asset pricing divide economics from finance (Marsh, 1984; Summers, 1985). Ultimately, 'perfect barter or frictionless models appeal [to economists] because it was thought that the successful empirical application of these frictionless models in financial economics could be repeated in monetary economics' (Rogers, 2006b, p. 3, n.

1).

The concept of the stochastic discount factor itself gives a generalised asset-pricing formula of the form:

$$p_t = \mathbb{E}_t[m_{t+1}x_{t+1}]. \tag{6.9}$$

Where m_{t+1} is any stochastic discount factor providing the state-contingent value of x_{t+1}, which represents asset dividends, cash flows or quasi rents (Cochrane, 2005b).

Note that, for the generalised formula, neither Rational Expectations nor complete markets are assumed; equation (6.9) is sufficient to establish asset prices, either in equilibrium or as individual measures of value. In addition, all returns are such that:

$$\mathbb{E}_t[m_{t+1}R_{t+1}] = 1, \tag{6.10}$$

where R_{t+1}, the asset's gross return, is defined as:

$$R_{t+1} \equiv \frac{x_{t+1}}{p_t}. \tag{6.11}$$

Every asset-pricing model implies an expression of the form of equation (6.9) and equation (6.10). The content of the model is in the interpretation of the elements.

It is a 'classic issue in finance' (Cochrane, 2005b, p. 15) that idiosyncratic risk does not warrant an expected return premium, since it is risk that can be diversified by, for instance, having a portfolio of equities rather than one single stock. If information is symmetric, so that all agents are equally well informed, then asset-specific risks can be diversified (O'Hara, 2003). Then only the risk that cannot be diversified, known as systematic risk, will have an expected return above the riskless rate of return. The asset's expected excess return is based on its sensitivity to the overall market, thus forming the basis of the Capital Asset Pricing Model (CAPM) (Lintner, 1965; Sharpe, 1964). It is worth noting that liquidity itself is similarly diversifiable because portfolios can be divided among various illiquid assets to reduce the risk that any one asset will not be liquid when necessary.

That returns are related only to systematic risk can be seen by decomposing equation (6.10) using the definition of covariance: $cov(m, R) \equiv \mathbb{E}[mR] - \mathbb{E}[m]\mathbb{E}[R]$ (Cochrane, 2005b, p. 14). Dropping sub-scripts, we have:

$$\mathbb{E}[m]\mathbb{E}[R] + cov(m, R) = 1. \tag{6.12}$$

Then, by introducing the gross risk-free return, $R^f \equiv 1/\,\mathbb{E}[m]$, we have an expression for the expected excess return:

$$\mathbb{E}[R^e] \equiv \mathbb{E}[R] - R^f = -R^f cov(m, R). \tag{6.13}$$

Interpreted through the lens of the consumption-based asset-pricing model, an asset's expected excess return is lower (and its price is higher) if it enables more consumption when the marginal utility of consumption is higher, that is, if the asset has a higher payoff in economic downturns.

The assumption of a risk-free rate, or 'common pure rate of interest' (Sharpe, 1964, p. 433), takes on vital importance and removes the need for a liquid asset. The risk-free asset, in a world of perfect market liquidity, dominates any other non-interest earning asset, such as money. In the Arrow-Debreu context, the risk-free asset can be constructed from contingent claims that span all future outcomes (Cochrane, 2005b, p. 51). In a Wicksellian model, the risk-free rate is equated to the marginal productivity of capital and the natural rate of interest (Rogers, 2007, p. 139). In presentations of the CAPM, for example, prices of risky assets adjust to reach a stock equilibrium, but the risk-free rate is exogenously given (Admati, 1985, p. 631; Lintner, 1965, p. 15) or implicitly the result of a loanable-funds flow equilibrium (Sharpe, 1964).

The risk-free rate also exemplifies the difficulties in moving from relative to absolute value. In a relative value framework, an asset's value can be implied from other tradable assets that have a similar risk profile. A financial options contract is valued by replicating its payoff by means of the underlying asset (Black & Scholes, 1973). When all risks are eliminated by replication then the investment that remains earns the 'risk-free' rate of return—an abstract construction representing the theoretical state of a perfectly known outcome.

Evidence suggests that discount factors vary through time and expected returns are, to a certain extent, predictable (Fama & French, 1988, 1989); expected returns do not simply follow a 'random walk' (Brown, 2011; Fama, 1991). Shiller (1981, p. 421) highlights the difficulties arising from the assumption that discount factors are constant. For if they are, sudden movements in stock markets can only be due to new information about future dividends, which are not volatile enough to explain observed stock price volatility. Asset prices are more volatile than their 'underlying fundamentals' would seem to imply

(Cochrane, 1992). 'Simple regressions of real ex post stock returns on lagged dividend yields find that the null hypothesis that the real ex ante rate is constant can be rejected at almost any level of confidence' (Summers, 1985, p. 635). Discount factors rise in economic downturns, which suggests that agents want higher returns when conditions are bad or uncertain (Crotty, 2011, p. 146) in such a way that 'the premium [on long securities] compensates for exposure to discount-rate shocks' (Fama & French, 1989, p. 24).

Furthermore, empirical studies show that 'higher or lower stock prices do not signal higher or lower subsequent dividends' (Cochrane, 2017a, p. 947), instead forecast excess returns rise when dividend/price ratios are low. 'High prices reflect low risk premia, lower expected *excess* returns' (Cochrane, 2005b, p. 400, original emphasis). Since 'asset price fluctuations are all about variation in risk premiums', it follows that 'recessions are driven by varying risk premiums and risk aversion' and not by variations in risk-free 'interest rates and intertemporal substitution of present for future consumption' (Cochrane, 2017a, p. 978). It should be noted, however, that empirical studies of expected returns focus on excess returns, a focus that influences their conclusions. Excess returns are founded on the concept of the risk-free rate, which is itself theory-laden. Consequently, studies that interpret empirical 'facts' as theory-neutral accidentally adopt a 'naive objectivist' approach (Sayer, 1992, p. 45; see Section 2.2).

The difficulty with a non-constant real rate of interest in theory can be highlighted by placing the asset-pricing formula into a Wicksellian model with production technology displaying stochastic constant returns to scale. In a single consumption-good model, consumption and investment are perfect substitutes and a temporarily high output of consumption goods will be transformed into capital because saving is investment (Marsh, 1984, p. 15). The profile of consumption and asset prices both remain unaffected because asset returns are determined by the marginal productivity of capital. The empirical variation in asset returns cannot be explained. Only in a model with non-storable goods can asset prices vary dramatically with variations in output. Then, asset prices are affected by the failure in aggregate for agents to 'save' excess output or conversely exchange assets for consumption goods when output is low (Lucas, 1978; Marsh, 1984).

Allowing the consumption-based discount factor to vary enough over time to explain observed asset-price variability amounts to adding an extra term y_{t+1} to equation (6.7) so

that[38]

$$m_{t+1} \equiv \beta \frac{u'(c_{t+1})}{u'(c_t)} y_{t+1}. \tag{6.14}$$

The introduction of an extra time-varying element means that discount factors can always be found to satisfy equation (6.10), thereby rendering it un-testable (Shiller, 1981, p. 430). Since 'distorting marginal utility is the same thing as distorting probabilities' (Cochrane, 2017a, p. 949), discount factors, or expected returns, are not directly observable. As seen in this section, insights cannot be obtained using the asset-pricing model within a positivist methodology, they are part of the 'deep stratum' (see Section 2.3).

The concepts of the REH and the EMH can be isolated by reference to the generalised asset-pricing formula, equation (6.9), or its equivalent $\mathbb{E}[m_{t+1}\, R_{t+1}|\mathcal{F}_t] = 1$. The EMH states that all available information (past and present) contained in $\mathcal{F}_t$ is used in the formation of the expectation $\mathbb{E}[.\,|\mathcal{F}_t]$.[39] It is because it makes no claims about the price of risk conveyed by the stochastic discount factor m_{t+1} that problems exist in testing the EMH. As Fama (1991, p. 1575) points out, 'market efficiency per se is not testable'. The empirical testing of the EMH cannot be separated from the researcher's model for asset pricing, since to determine whether 'information is properly reflected in prices', some model of 'properly' is necessary (Fama, 1991, p. 1576). Fama calls this the 'joint-hypothesis' problem and failures of empirical tests can result simply from a 'bad model of market equilibrium'.

Only with the addition of the REH do agents use the objective, or model-consistent, probability distribution of the future cash flows x_{t+1} to establish the expectation $\mathbb{E}_t[.]$ (Weitzman, 2007). The REH is required to establish unique prices in a competitive equilibrium, but the 'EMH does not imply that prices are set in some kind of competitive market equilibrium' (Brown, 2011, p. 82). The EMH does not imply the REH, but the REH implies the EMH. The assumption of REH is not essential; subjective probabilities, that vary by agent, could be used to form the expectations of cash flows (Cochrane, 2005b, p. 54). The general applicability of the formula allows its use in the evaluation of Keynes's

[38] See Cochrane (2017a, p. 948).

[39] The information set $\mathcal{F}_t$ is equivalent to Keynes's (1921) concept of hypothesis or evidence. See Section 8.4.

theory of liquidity preference in Section 8.7.

The importance of distinguishing the information set and the pricing model in the asset-pricing expression is that, ultimately, the probabilities used to determine asset prices are not based on the 'correct expectation of the distribution of future cash flows associated with each security', as Crotty (2011, p. 136) asserts. Asset prices are based on risk-neutral probabilities. The distinction between actuarial and risk-neutral probabilities should be recognised. Although risk-neutral probabilities are related to actuarial or real-world probabilities, they are not same.[40] It is not necessary to know the 'true' actuarial probabilities to estimate the risk-neutral equivalents. Real-world probabilities are unknowable and are subject to fundamental uncertainty (Crotty, 2011; Hayes, 2006), but risk-neutral probabilities are subjective and differ for each individual. Because they are a combination of subjective judgements of utility and likelihood, their estimation suffers from the joint-hypothesis problem. This hypothesis-testing problem is generalised as the Durhem-Quine thesis, which states that it is impossible to isolate a single hypothesis for testing (Boumans & Davis, 2010; Elsner et al., 2014).

The foregoing discussion affects the interpretation of the observation that expected returns (and hence the stochastic discount factor or pricing kernel) vary over time. In the view of Fama (1991, p. 1577), this variation is likely to be 'real and rational'.

> The predictability of stock returns from dividend yields…is not in itself evidence for or against market efficiency. In an efficient market, the forecast power of D/P [dividend yield] says that prices are high relative to dividends when discount rates and expected returns are low, and vice versa. On the other hand, in a world of irrational bubbles, low D/P signals irrationally high stock prices that will move predictably back toward fundamental values. To judge whether the forecast power of dividend yields is the result of rational variation in expected returns or irrational bubbles, other information must be used. As always, even with such information, the issue is ambiguous. (Fama, 1991, p. 1583)

This predictability covers all asset classes. The evidence regarding expected returns is that

[40] Often the real-world measure is denoted by $\mathbb{P}$, and the risk-neutral measure as $\mathbb{Q}$. Technically they are 'equivalent probability measures': for any set, A, $\mathbb{P}(A) > 0$ if and only if $\mathbb{Q}(A) > 0$. In other words, any outcome possible by $\mathbb{P}$ is also possible by $\mathbb{Q}$, and vice versa. See Harrison and Kreps (1979, p. 383).

there exists a ‘common premium for maturity risks’ (Fama, 1991, p. 1584) or ‘common variation in expected returns’ (Fama, 1991, p. 1585). Fama (1991, p. 1585) also points the way forward, his ‘view is that we should deepen the search for links between time varying expected returns and business conditions, as well as for tests of whether the links conform to common sense and the predictions of asset pricing models.’ Cochrane (2011b, p. 1071) would also move the focus from the incorporation of information to ‘discount rates, expected returns, risk bearing, risk sharing, and risk premiums’, since the ‘[EMH] basically won, and we moved on. When we see information, it is quickly incorporated into asset prices’ (Cochrane, 2011b, p. 1072). Whether or not markets are ‘right’, it is sufficient to recognise that markets can incorporate new information without actual trading and that expected returns vary over time. This variability in expected returns common over all assets introduces the uncertainty in the future rate of interest underlying Keynes’s theory of liquidity preference, to be discussed in Chapter 8. These empirical tendencies also influence later discussions about the importance of a variety of opinion in the market in Section 8.5, and whether it is the flow of risk or the flow of funds that affects asset prices in Section 8.6.

6.7 Summary

Liquidity has an uncomfortable position in Real Analysis. The Arrow-Debreu complete-markets model of neo-Walrasian general equilibrium equates to a state of pure exchange and has no essential need for money or money flows. Neo-Walrasian theory, by abstracting from liquidity scarcity, assumes perfect market liquidity and hence liquidity abundance. All other forms of liquidity are absent and money has no fundamental place in the analysis. Closer inspection also reveals that no financial securities of any sort are necessary. Rogers (2014) highlights the conceptual problem: attempting to analyse liquidity, by any sensible definition of the term, is simply impossible in a model without money. Similarly, the Fisher Effect, whereby changes in expected inflation have a one-for-one effect on nominal interest rates, is only justifiable in a non-monetary setting.

Wicksellian theory, on the other hand, recognises liquidity scarcity, the credit nature of money, the importance of the banking system, and a market rate of interest that clears the money market in equilibrium. Since the theory is anchored to the concept of a natural rate of interest determined entirely by the real economy it also follows in the tradition of Real Analysis and has a limited role for either liquidity or money. Furthermore, the natural rate

of interest itself can only be justified theoretically in a single-commodity model, which also precludes any essential role for a monetary asset by allowing perfect technological liquidity to obviate the need for any other form of liquidity. Savings in a single-commodity world can directly 'finance' investment. Consequently, neither a banking system nor money is required. The Wicksellian notion of a natural rate of interest determined entirely in the real economy stems from confusion between money and saving and results in theoretical inconsistencies that also call into question the value of its analytical insights. An analytical system built on Wicksellian foundations cannot provide a fundamental explanation of liquidity.

General equilibrium asset-pricing methodologies are based on a number of problematic assumptions. The most relevant to this study is the exclusion of liquidity since it is observed to have significant effects. That said, much of modern asset-pricing theory is used in practice and cannot be simply rejected. The focus on risk as the basis of pricing, rather than the supply and demand of assets, is more in tune with the intermediated nature of finance and the credit theory of money itself. When financial derivatives can be created to refer to component aspects of asset risk, the supply and demand of specific assets cannot remain the focus. The issue to be tackled is that risk-pricing paradigm is supported by the assumption that all risks can be traded in any amount. It must be recognised that it operates in an idealised liquidity-abundant world and that, if possible, liquidity must be incorporated.

The crux of the problem is that general equilibrium asset pricing has no fundamental place for liquidity, even though its effects are recognised by the associated finance literature. Instead, something called 'liquidity' can only be incorporated into these pure-exchange models by the addition of arbitrary frictions that only temporarily prevent the ultimate general equilibrium. Although the treatment of uncertainty in the literature is side-lined by the use of Rational Expectations due to its tractability, aspects of the risk-neutral approach are useful. In the next chapter, we consider various attempts to incorporate liquidity and money into general equilibrium theory.

7 Liquidity and Money in General Equilibrium

7.1 Introduction

Attempts to incorporate a monetary asset into either Wicksellian or the neo-Walrasian general equilibrium theories take many forms. Most can be classified as *ad hoc*, in that they do not present a coherent explanation of the existence of money; they simply recognise that money is present in the real world and make a pragmatic attempt to rectify this absence from the analytical world. The attempt to explain liquidity from within a framework of perfect market liquidity, leads to the opposite: how to explain illiquidity or the lack of liquidity. The task reduces to looking for imperfections along the lines of 'participation costs, transaction costs, asymmetric information, imperfect competition, funding constraints, and search' (Vayanos & Wang, 2011, p. 221). By concentrating on market liquidity, which eclipses all other forms of liquidity, general equilibrium models are unable to analyse the various forms of liquidity identified in Chapter 5. This chapter considers the range of attempts to add liquidity and money to models where they have no fundamental place.

7.2 Utility of Money

Money can be added as a commodity-like asset issued by government fiat. Immediately, problems arise with its valuation. If, for greater realism, the fiat asset is also used as the *numéraire* then there are additional problems with price-level determinacy. For example, in finite time-horizon models the terminal value of money must be zero, since it provides no consumption utility, and by backwards induction, its value must be zero in all previous periods. In infinite lifespan models, the perception that money is a zero-coupon perpetual bond causes similar valuation problems. To give money a formal value, an *ad hoc* appeal must be made to its unspecified usefulness in facilitating trade and reducing transaction costs (Woodford, 2003, p. 102). This usefulness can then justify the explicit inclusion of money balances in the representative agent's utility function.

Once money is included directly in the agent's utility, the standard neoclassical optimising techniques can be applied, so that the marginal utility of money can be used to establish the

demand for money. To prevent any form of money illusion[41] the money balances in the utility function must be adjusted by the price level or the value of money itself. The stock of money must also be homogeneous and fixed by some external agent, such as a central bank. The scarcity of liquidity must be maintained by the strict separation of credit from money-proper. For example, Kiyotaki and Wright (1989) assume a 'continuum of infinitely lived agents' to ensure that 'there is no credit since a given pair will meet again with probability zero' (p. 931). This set-up achieves 'sufficiently well-defined demand and supply functions to give the unit in which [the fiat money] is measured determinate prices in terms of other goods' (Fama, 1980, p. 50).

Notwithstanding the lack of explicit justification for the inclusion of money, the approach presents some logical problems, the first of which is easily disposed of but needs a mention nonetheless. The 'circularity charge' alleges that the marginal utility of money cannot determine the value of money since the price level must first be known before the marginal value of an extra unit of money can be established. The '[circularity] charge originates in a basic misunderstanding of the theory of price determination' (Patinkin, 1965, p. 114). The key is to differentiate between individual experiments, where hypothetical quantities are determined for given prices, and market experiments, where prices are determined by the interplay of these individual demands and supplies to establish a market equilibrium (Patinkin, 1965, p. 12). Along these lines, the utility of monetary balances is the result of an individual experiment conducted at varying prices; the value of money itself is determined in a market experiment based on the interplay of these individual experiments (Niehans, 1978, p. 13; Patinkin, 1965, p. 116).

The more challenging issue in applying utility theory to money is that 'the utility of money has [both] a flow aspect and a stock aspect' (Niehans, 1978, p. 14). The former concerns the purchasing power of income in monetary form, its 'subjective exchange value' (Schumpeter, 1954, p. 1089), where the utility of extra units of monetary income is the value of goods that can be purchased. The flow value presents no difficulty since 'the proposition that money's utility derives solely from its exchange value is indisputable' (Chick, 1983, p. 303). The problem arises in determining the stock aspect of utility, that is,

[41] An individual is free of money illusion if his excess-demand functions depend solely on relative prices and real wealth (Patinkin, 1965, p. 22).

the value of money held. The problem with balances of purely fiat money is that they do not improve anyone's wellbeing. It is here that deductive 'as if' reasoning enters. The security of holding real balances of money, in an environment where future payments are at uncertain times, is simply assumed to provide utility similar to that of consumption goods (Patinkin, 1965, p. 80). In a sense, 'these approaches must presuppose what they set out to explain' (Ingham, 2000, p. 20) because, although the utility of consumption goods is well understood and concerns individual tastes, the utility of money balances is not, and is merely a proxy for unexplained 'frictions in the exchange system' (Niehans, 1978, p, 16). Money is deemed to provide an unspecified and unexplained liquidity service, which is nevertheless valued by a single, representative agent.

A more explicit attempt to include money in neoclassical models, by assuming that money reduces unspecified transactions costs, leads to the cash-in-advance model (Clower, 1967; Lucas, 1990) where all purchases must be made in a single, state-issued money. Effectively, a medium of exchange is introduced into a neo-Walrasian general equilibrium model by imposing the equation of exchange, equation (3.1). The demand for money is established because the representative agent must have sufficient money balances to implement his proposed consumption plans. Under some conditions, however, this restriction prevents otherwise-optimal trades. Instead, money operates as a 'finance constraint', and 'cannot be interpreted as a demand and supply relationship in neo-Walrasian theory' (Rogers, 1989, p. 98). The cash-in-advance constraint designed to integrate money into a fundamentally moneyless model is welfare reducing and operates as an impediment to trade or by increasing the volatility of asset prices. The lack of explanatory content and counter-intuitive nature of these models leads to post-GFC exasperation: 'The irony of modelling liquidity by imposing money as a constraint on trade was lost on the profession' (Buiter, 2009, p. 1).

Overlapping-generations (OLG) models have also been used to provide an avenue for the existence of an, otherwise worthless, fiat money asset (Samuelson, 1958). In these models, agents generally 'live' for two periods. In the first, they are 'young' and productive and can earn the money asset. In the second, they are 'old' and unproductive; they must live off their saved money. Overlapping generations of young and old agents coexist and this allows the transfer of money assets from the old to the young in each period. The acceptance of money, however, is only ensured by eliminating the possibility of saving in land or other

durable goods. Restricted to perishable endowment goods during their productive period, young agents have no way to save or invest for their 'retirement' period. The introduction of paper 'money' allows for this, but the result is that saving and investment are physically identical, not just an accounting entity. OLG models do not explain 'money' because the object representing 'money' is not money at all, and would be dominated by any other interest-earning asset (Rogers, 1989, p. 46-47).

Despite their continued use, the only conclusion regarding cash-in-advance and overlapping-generations models is that 'they have no hope of explaining endogenously either the nature of money or the development of monetary exchange' (Kiyotaki & Wright, 1989, p. 928, n. 1). Similarly, the observation that the medium of exchange is subject to evolution in its objective form is a basis for rejecting the cash-in-advance technique or putting money in the utility or production functions (Kiyotaki & Wright, 1989, p. 951). This problem is compounded when the hierarchical basis of liquidity and money, identified in Part II, is considered. These techniques neither capture the tendency of the banking system to evolve (Minsky, 1957) nor cover the 'wider structure of liquidity' beyond the supply of money (Radcliffe Committee, 1959, p. 132). Adding a commodity-like money asset to general equilibrium models does not provide a fundamental treatment of liquidity or money.

7.3 Financial Frictions

Rather than adding money directly into general equilibrium models, the assumption of perfect market liquidity itself can be addressed. Attempts to include the effects of liquidity in the context of a pure-exchange general equilibrium model are made by adding *ad hoc* 'financial frictions'. These frictions provide an 'as if' element to introduce the market illiquidity and funding illiquidity observed during financial crises. The frictions can take various forms, such as inefficient use of technology, transaction costs, or imperfect information; all serve to create instances of incomplete markets (Quadrini, 2011). Without perfect market liquidity, some trades that would otherwise occur are prevented.

Just so, Cochrane (2014) outlines two 'stories' why financial crises affect the real economy. In the first, falling asset values prevent the banking system from playing its role as an intermediary between savers and lenders. A financial crisis is a form of financial friction, in which otherwise-profitable firms cannot get access to funds. In the second story, which

Cochrane calls ‘aggregate demand’, a flight to quality assets means higher government bond prices, lower equity prices and a subsequent fall in consumption and investment demand. Institutional investors, for some reason, sell assets temporarily undervalued by the crisis. In Cochrane’s opinion, the second story ‘can easily be managed with sufficient fiscal and monetary stimulus’ so that there is no need to adjust banking policy (Cochrane, 2014, p. 211).

Financial frictions do not address the conceptual problems inherent in the money-in-utility or cash-in-advance models. The moneyless neo-Walrasian model is kept intact; the frictions merely prevent the system from attaining its full, efficient equilibrium. At its core, each model assumes the existence of a ‘true’ general equilibrium price vector. The frictions create the need for enhanced, illiquidity-compensating returns for some agents, which are then identified as liquidity premiums. None of these additions creates a demand for a liquid store of value. Instead, they are restrictions on the success of the invisible medium of exchange and not a deep, causal explanation of liquidity itself. Consider following statement taken from a survey of ‘financial frictions’, directly after a recognition that ‘liquidity spirals’ and the dysfunction of a ‘credit crunch’ are major contributors to the financial sector’s instability:

> In a frictionless economy, funds are liquid and can flow to the most profitable project or to the person who values the funds most. Differences in productivity, patience, risk aversion or optimism determine fund flows, but for the aggregate output only the total capital and labor matter. Productive agents hold most of the productive capital and issue claims to less productive individuals. In other words, in a setting without financial frictions it is not important whether funds are in the hands of productive or less productive agents and the economy can be studied with a single representative agent in mind. In contrast, with financial frictions, liquidity considerations become important and the wealth distribution matters. (Brunnermeier et al., 2013, p. 4)

The use of the term ‘funds’ and the phrase ‘funds are liquid’ in the quotation above implies a monetary flow, but in the neo-Walrasian model there are no money flows, since all trades are direct via commodity futures contracts. The phrase ‘funds are liquid’ gives the impression that money is an aspect of the analysis, but the correct interpretation is that it is resources that are mobile, there is no movement of a medium of exchange. Financial

frictions build on this monetary masquerade by conjuring up an image of a 'monetary flow' that is somehow impeded by the financial system. The financial system cannot impede a non-existent flow, so the idea of financial frictions collapses (Buiter 2009; Rogers, 2006c, 2008a; Rogers, 2014).

A family of models (Holmström and Tirole, 1998; and Tirole, 2008) provides an illustrative attempt to analyse liquidity using liquidity-free general equilibrium theory. Tirole recognises up front that general equilibrium models cannot find a fundamental place for liquidity or money:

> But what is "liquidity"? Does liquidity matter and should governments and central banks do something about it? While trivial to a practionner [sic], these questions surprisingly are not so obvious to an economist trained in the general-equilibrium tradition. Intuitively, an industrial company or financial institution is short of liquidity when a) some (continuation or investment) spending decisions are worthwhile, and b) the firm somehow cannot manage to find the money to finance them. Classical (Arrow-Debreu, Modigliani-Miller) economic theory holds it that a) and b) are inconsistent: if refinancing or financing of new projects is desirable, so goes the argument, the firm can always issue claims on associated future profits, that investors will find sufficiently attractive to be willing to finance the outlay. According to this logic, firms have no reason to plan their liquidity (or for that matter to engage in risk management to avoid bad surprises in their liquidity position): they just can return to the capital market as needs arise. (Tirole, 2008, p. 54)

In response, Tirole (2008, p. 54) provides a 'conceptual framework for liquidity', to resolve the inconsistency in Arrow-Debreu models of not being able to find funding for profitable activities. The framework incorporates agency costs and costly refinancing unlike the 'classical economics world, in which firms could costlessly return to the capital market to raise funds when they need to' (Tirole, 2008, p. 56). Unlike in the Arrow-Debreu world, where no trading is conducted beyond the Walrasian auction, cash flow mismatches arise which must be funded by 'finance as you go' or 'liquidity hoarding' (Tirole, 2008, p. 55). Specifically, a loss, deemed a 'liquidity shock', in the second period of a three-period model must be covered to ensure the realisation of the investment gains available in the final period. External financing is unavailable, however, because there is a 'wedge between the

full value of the firm and the external value of the firm' that 'prevents it from financing all projects that have positive net present value' (Holmström & Tirole, 1998, p. 2), so the potential gains in the subsequent period do not sufficiently remunerate the new investors.

Rather than modelling a liquidity event, where funds are unavailable for otherwise profitable activities, the loss is a catastrophic event. The supposed 'liquidity shock' requires new funds, but not for rolling over existing funds to resolve a funding liquidity problem, or relying on market liquidity to sell an asset that has fallen in value. It is more akin to an insurance loss, where an income-generating asset has been destroyed and must be rebuilt to complete the partially produced output for sale. Liquidity is presented as some sort of 'agency problem'—an *ad hoc* explanation that has no theoretical content, a symptom of how general equilibrium models struggle to 'illustrate how disruptions in financial intermediation can induce a crisis that affects real activity' (Gertler & Kiyotaki, 2011, p. 549). It appears that, unless the 'refinancing' involves a loss, the Arrow-Debreu model cannot conceive of any reason for a failure to source funding.

For Tirole, the financial construction that solves the problem is a 'line of credit', but it actually takes the form of an insurance contract. This misunderstanding is evident in the description of a credit line:

> This is indeed the nature of a credit line: there would be no reason to contract in advance on a credit line if at date 1 the bank were always happy to provide the funds; it is precisely because lending is a money-losing operation at date 1 that it must be pre-arranged. (Tirole, 2008, p. 58)

The suggestion, and hence the basis for a fundamental explanation of liquidity, is that banks make 'money-losing' loans on the basis of pre-arranged credit lines. This is simply not true in practice. Credit lines are contingent on the credit-worthiness of the borrower, and both terms and availability are revised based on credit events like the one underlying the 'liquidity' in the model. Perpetuating loanable funds' confusion of saving and money, there is uncertainty regarding where the banks find the funding to satisfy the pre-arranged lines of credit (Tirole, 2008, p. 58). The commodity view of money underlies the question as to 'whether the intermediary can raise enough funds at date 1 to meet its (deterministic) credit obligations' (Holmström and Tirole, 1998, p. 18). With a more realist assessment of the banking system, this question would not be asked.

Instead, statements such as ‘credit lines are granted by intermediaries such as banks, which can make commitments, assuming of course that they have enough funds to meet later credit needs’ (Holmström and Tirole, 1998, p. 13) should be re-framed in terms of whether it has enough collateral to borrow these funds from another bank intermediary. In the absence of aggregate uncertainty, the liquidity requirement is deterministic and can be met by the private sector, whereas aggregate uncertainty requires government intervention (Holmström and Tirole, 1998, p. 15). In other words, government can issue safe assets that are used as the basis of outside liquidity. Mehrling (2012) emphasises collateral as key to keeping the liquidity of the future incarnation of the shadow banking system alive. Central banks cannot maintain the system by being only the lender of last resort, an issue considered further in Section 9.4.

As discussed in Section 6.4, in Real Analysis banks are merely financial intermediaries with no special significance and can therefore be ignored. Nevertheless, the Diamond-Dybvig model (Diamond & Dybvig, 1983), and its derivatives (Allen & Gale, 1998, 2000, 2004), is the workhorse model of bank runs and liquidity provision, and requires discussion and placement in its theoretical context. The model is briefly as follows: ‘Illiquidity’ originates from the different yields available from operating the productive technology over one or two periods in a three-period model. Interrupting the technology after one period reduces the yield compared with its operation over two periods. Although the model has ‘no transaction costs’, the ‘analysis would be the same if the asset were illiquid because of selling costs: one receives a low return if unexpectedly forced to “liquidate” early’ (Diamond & Dybvig, 1983, p. 403).

Demand for ‘liquidity’ arises because, in the second period, agents discover their personal consumption ‘type’ or ‘liquidity preference’ (Allen & Gale, 2004, p. 1016), that is, in which of the second or third periods they prefer to consume the single homogeneous good. The consumption type is only known to the consumer, and is not publicly available, thus preventing a full neo-Walrasian equilibrium and ‘optimal risk sharing’. There can be no ‘complete market of Arrow-Debreu state-contingent claims, because this market would require claims that depend on the nonverifiable private information’ (Diamond & Dybvig, 1983, p. 407). The model is based on a financial friction arising from incomplete information.

In the model, bank deposits can mitigate the problem of incomplete markets, but at the risk

of bank runs. 'It is precisely the "transformation" of illiquid assets into liquid assets that is responsible both for the liquidity service provided by banks and for their susceptibility to runs' (Diamond & Dybvig, 1983, p. 409). Consumers can 'deposit' their first period endowment of the consumption good into the bank. This deposit returns r_1 units of the consumption good on demand in the second period, regardless of the consumer's type. In this case, customers are served sequentially 'until the bank runs out of assets' (Diamond & Dybvig, 1983, p. 408). The bank is susceptible to a bank run when $r_1 > 1$; that is, when the return is greater than the 'liquidating' value of the technology. If the value of the bank deposit in the second period were the same as the liquidating value of the technology then there would be no bank run, but a 'demand deposit contract which is not subject to runs provides no liquidity services' (Diamond & Dybvig, 1983, p. 409).

Lending is conducted in terms of a single consumable or investible good, and the 'liquidity' constraint is due to the time profile of the available technology. The form of 'liquidity' adopted is technological and arises from a 'real' production constraint. To isolate the effects of 'liquidity', as distinct from 'risk', only the uncertainty regarding the timing of consumption is allowed to affect the equilibrium. 'There is a trade-off between [technological] liquidity and returns: long-term investments have higher returns but take longer to mature (are less liquid)' (Allen & Gale, 2004, p. 1016). As such, the liquidity premium is assumed rather than explained. There is no risk-return trade-off: production outcomes (i.e. 'asset returns') for the two production techniques are deterministic. The external, technological specification of asset returns based on the existence of a single consumption good gives the model a Wicksellian pedigree.

Diamond and Dybvig (1983, p. 404) interpret the results as having 'far-reaching policy implications, because they imply that the real damage from bank runs is primarily from the direct damage occurring when recalling loans interrupts production.' In the strict sense, banks cannot recall loans in this fashion, but the phenomenon could be interpreted as the losses made by the bank when attempting to sell illiquid assets. As such, the model reflects some of the difficulties faced by banks in maintaining convertibility to state money. There are, however, some ontological problems.

First, the bank 'deposits' are contracts to provide a 'fixed amount of consumption' (Allen & Gale, 2004, p. 1017), there are no monetary aspects to the model, and as such it is a non-monetary form of analysis. The 'bank' is unrecognisable as a money-creating entity that

exchanges and clears credit. The bank must receive deposits, or prior saving, before engaging in investment projects. It is standard within the financial frictions literature to 'assume that households deposit funds in financial intermediaries that in turn lend funds to nonfinancial firms' (Gertler & Kiyotaki, 2011, p. 550), thereby disallowing any asset-price feedback between credit creation and existing assets (Turner, 2013, p. 30). The model confuses money and saving and suffers from all the problems identified with loanable funds theory.

Second, and related to the first problem, the 'bank' attempts to optimise the production of the consumption good in the second and third periods, subject to the technological trade-offs and the randomness of consumption preferences. Hence the 'bank' is, in fact, a planner, whose role it is to maximise the expected utility of consumers (the 'depositors'). Its portfolio of long- and short-term investments can differ from the demands of the depositors and this misalignment in the flow of the consumption good is the source of the bank run.

The Diamond-Dybvig model is used to interpret the GFC, which has been likened to a traditional bank run in the shadow banking system (Gorton & Metrick, 2012; Lucas, 2014; Lucas & Stokey, 2011). Repo agreements replaced bank deposits, but without deposit insurance or the protection of a lender of last resort. Lucas and Stokey (2011) apply the model to analyse this 'run on repo'. They conclude that repo collateral plays the role of deposit insurance (Lucas & Stokey, 2011, p. 12). Michell (2017) highlights the analytical and empirical flaws in this argument. First, it is essentially a category error to compare the banking system, which creates money, to the repo market, which cannot. Second, the bank run argument fails to recognise the importance of asset-backed commercial paper (ABCP) as the primary source of financing for the shadow banking system. Only once the shadow banking system had moved through its first phase of failure, in which the ABCP market had ceased to function, did the system resort to repo financing (Mehrling, 2011). The shadow banking system relied on repo financing during the second phase, before moving to the third phase in which off-balance-sheet vehicles were brought back onto bank balance sheets.

Finally, it must be noted that the Diamond-Dybvig model covers only an individual bank-run; a system-wide bank run is a 'coincidence' (Allen & Gale, 2004, p. 1019). In fact, unless all bank depositors attempt to convert their holding into state money, an individual bank run cannot have systematic effects, therefore it is 'hard to invent persuasive scenarios in

which runs now could threaten the banking system as a whole' (Moore, 1988, p. 37, n. 19). Thus:

> A run on an individual institution is not a crisis, however. To be a crisis, the run has to affect the financial system and, ultimately, the real economy. We need to understand what makes for "contagion" or a "systemic" run. (Cochrane, 2014, p. 206)

In Real Analysis, financial crises are ideally related to changes in the fundamentals of 'tastes, endowments, or production possibilities' (Cass & Shell, 1983, p. 194, n. 1), so that contagion derives from 'real shocks and real linkages' (Allen & Gale, 2000, p. 3). These shocks represent intrinsic uncertainty 'caused by stochastic fluctuations in the primitives or fundamentals of the economy' (Allen & Gale, 2004, p. 1017). Faced only with intrinsic uncertainty, complete markets are able to achieve full risk sharing and establish a 'fundamental equilibrium', so that a 'crisis cannot occur in a fundamental equilibrium in the absence of exogenous shocks to fundamentals, such as asset returns or liquidity demands' (Allen & Gale, 2004, p. 1018).

Exogenous shocks, or 'sunspots', however, allow for extrinsic uncertainty which 'by definition has no effect on the fundamentals of the economy' (Allen & Gale, 2004, p. 1018). Extrinsic uncertainty cannot be shared because there are 'no markets for Arrow securities contingent on the sunspot variable, so financial institutions cannot insure themselves against asset-price fluctuations associated with the sunspot variable' (Allen & Gale, 2004, p. 1020). Consumers are excluded from the capital market and are unable to construct contracts to 'implement the first-best allocation of risk' (Allen & Gale, 2004, p. 1023). As such, the models are open to the objection that they 'depend on easily cured ignorance about the current state of the world and on inability to create simple securities' (Black, 1987, p. 147, n. 7). Nevertheless, Lucas and Stokey (2011, p. 10) interpret a sunspot equilibrium as representing an outcome that occurs due a self-reinforcing belief feedback loop where 'what you want to do depends on what you think others will do.' The sunspot introduces a systemic element missing from the Diamond-Dybvig model and overcomes the weaknesses in a 'domino' view of financial crises versus a 'contagion' view (Cochrane, 2014). Rather than losses at individual financial institutions causing subsequent institutions to topple like dominoes, the GFC showed that systemic effects on asset prices are able to cause all institutions trouble at once.

Using the Diamond-Dybvig model to evaluate bank runs also leads Cochrane (2014) to recommend that banks be prevented from issuing 'run-prone securities'. Instead of deposits, banks should fund themselves with 100% equity[42], since equity holders cannot force repayment; they can only sell their equity. The ontological underpinning of this recommendation is the commodity view of money, whereby there is a fixed stock of money in circulation such that 'system as a whole promises more cash than is available' (Cochrane, 2014, p. 206). In this worldview, liquidity cannot be differentiated from solvency, and a failure to repay in the money commodity is insolvency. It follows that only asset sales represent liquidity, and, since banks only lend money they receive on deposit, they should be able to conduct asset sales to recover this money.

Attempts to prevent the possibility of bank runs by increasing the level of equity funding downplay liquidity risk to emphasise solvency risks. From this viewpoint, liquidity risk arises because of insolvency risk, for 'when banks take risks with the funds that they obtain by "producing debt," this risk can create solvency problems that would threaten the very liquidity' that banks create (Admati & Hellwig, 2013, p. 295). No recognition is given to the fact that insolvency is not a liquidity problem for firms with long-term liabilities, only when debt is short term. It is because bank liabilities are generally short-term, and hence liquid, that there is a problem. If not, solvency would not be an issue. Liquidity is the first-order problem, solvency is a subsidiary issue. Although the 'short-term debt that a bank produces may therefore not be very money-like if the bank does not have much equity' (Admati & Hellwig, 2013, p. 156), equity alone cannot save a bank from illiquidity.

Solvency-based analyses of bank liquidity ignore several issues with the asset side of the balance sheet. Descriptions of a 'traditional bank' (Admati & Hellwig, 2013, Chapter 4), with long-term lending funded short-term, are not accurate representations of traditional banking (see Section 4.4). If the assets are also short-term self-liquidating bills, in the tradition of English banking as per the recommendations of Bagehot and Dunbar, then the

[42] The 100% equity plan is very different from the 100% reserves or 'Chicago' plan most associated with Milton Friedman. The latter refers to having a one-for-one backing of state money for all bank money issued, and is more akin to the ideas of the Currency School (see Section 4.3) and the commodity view of money. The implications of the 100% equity plan are the reverse, since it would mean that banks would not be able to issue any debt or create money at all.

fire-sale issues arising from maturity transformation are less relevant. Using short-term funding for long-term assets, and thereby relying on market liquidity, leads to asset-price dislocations during a period of funding stress. Nevertheless, this dislocation is predominantly a liquidity event, and insolvency during these periods is primarily due to illiquidity. Proposals for the central bank to act as a 'dealer of last resort' are entirely tailored to the situation of long-term borrowing with short-term funding (Mehrling, 2011). In this case, we must recognise the importance of collateral and its implications for solvency.

Also downplayed by Admati and Hellwig is the moneyness of bank money, with the view that a 'dollar in a bank account is the debt of the bank. A dollar in cash is nobody's debt', since 'the issue of [government] banknotes does not commit the government to anything' (Admati & Hellwig, 2013, p. 294). By this reasoning, the gold standard, whereby central banks promise to pay gold of a fixed weight—a gold debt owed by the central bank—cannot be a monetary system since there is nothing answering to this definition of 'cash'. As seen in Section 5.7, the Modigliani-Miller theorem does not apply to firms that finance themselves in the money market. It may be that the equity funding of banks should be greater than pre-GFC levels, but recommendations based on an analytical framework inconsistent with fundamental ontological aspects of the banking system are of dubious merit.

7.4 Credit

Arrow-Debreu models, in addition to lacking any form of recognisable money, also lack the concept of credit risk. Inasmuch as the concept of monetary settlement is relevant, all transactions are settled at the beginning of time. Credit risk, in the sense that payment or final settlement is not forthcoming, simply does not feature. The Arrow-Debreu model determines pre-settled delivery of the contracted commodity. As such, there is no risk of default in money terms; there is no credit risk. From the ontological viewpoint of this study—that money is a form of credit—this omission is significant. The Arrow-Debreu securities are promises to deliver a commodity, not to pay money. The only risk is non-delivery of the commodity at the maturity of the futures contract, but even this risk can be mitigated or shared.

It can be imagined that all Arrow-Debreu markets clear by a form of 'trade credit' (Lucas,

1984, p. 21) that never defaults and is always accepted. Consequently, no agent has any need to maintain a reserve of better-quality credits, as is demonstrated by the hierarchy of money. The fundamental problem with adding money to an Arrow-Debreu model is not just that it appears as a finance constraint, or that it is unnecessary; credit money, as a medium of exchange, implicitly exists but, because of the lack of uncertainty and intertemporal trading, it ceases to exist in equilibrium. Instead, all banks work perfectly, all credits clear, and final settlement is achieved across all agents. Barter exchange, in which commodities are traded with a time lag, conceptually involves promise to pay, but the payment is made with other commodities. A 'credit' economy is possible with barter, but a monetary economy only appears when payment is made with the credit of others. The banking system, which enables settlement by means of the credits of others, is essential to a theory of money.

Once an allowance is made for potential delivery failures, however, the equilibrium becomes game-theoretic. Each agent must attempt to evaluate interdependent strategies of own-default and other-agent default. Default and bankruptcy can be 'thought to be incompatible with equilibrium analysis' and 'incompatible with perfect competition, since rational agents would have to calculate the probabilities of their loans defaulting, and conjecture how these probabilities would change as they increased their loans' (Geanakoplos, 1990, p. 32). The implications for a model where the basis of money is credit are significant. Within neo-Walrasian theory, default risks must be assessed during the time-0 Walrasian auction. No strategy can involve holding the credits of other agents—the pre-condition for a monetary asset. The characteristics of the resulting equilibrium are likely to differ from that under complete markets, but the process is no different.

For Goodhart (2005b, p. 195) 'the possibility of default is central to the analysis of any monetary world'. The absence of default risk for Goodhart is expressed via the transversality condition, equation (6.8), which he interprets as meaning that the representative agent '*always* manages, in due course, to pay her debts' (2005b, p. 194, original emphasis). Although the transversality condition is often interpreted as the assumption that agents never default (see Goodhart, 2009; Espinoza, Goodhart, & Tsomocos, 2009), in neo-Walrasian theory, as we have seen above, default is not a concept that exists. Instead, the transversality condition rules out endless debt accumulation and price bubbles (Cochrane, 1992, p. 247), not debt default.

The attempt to adjust asset prices for credit and liquidity risk from within the risk-neutral complete-markets world has created much confusion. From within this incoherent framework, the post-GFC emphasis on credit and liquidity has produced a suite of 'valuation adjustments', known collectively as 'XVA', to incorporate the elements missing from the complete, risk-neutral theory. Examples are the credit valuation adjustment (CVA), the funding valuation adjustment (FVA), the debt valuation adjustment (DVA) and the capital valuation adjustment (KVA) (Albanese, Caenazzo, & Crepey, 2016). Each adjustment is motivated by the omission of credit or default risk from the classic Arrow-Debreu model, and attempts to extend asset-pricing models to incorporate credit or liquidity risk by means of funding costs and credit or liquidity premiums. Nevertheless, all leave the complete markets framework intact and assume the existence of a risk-free rate. For instance, the 'FVA debate' (Castagna, 2012; Hull & White, 2012) sets risk-neutral theoreticians against market practitioners over the validity of adjusting asset prices to account for difficulties in accessing funding during a crisis. The theoreticians apply strict, but misguided, complete markets logic to the problem of liquidity and money to reject the effect of funding on prices. By contrast, the practitioners are theoretically adrift, but, by necessity, have the intuitive and pragmatic approach required to solve a real and pressing asset-pricing problem.

The lengthy debate between theoreticians and practitioners concerning the validity of the valuation adjustments, and how to incorporate them into the standard asset-pricing techniques, serves to highlight the unrecognised difficulty in attempting to incorporate failures of monetary payment or access to money into a framework that excludes money. Neither side of the debate has explicitly recognised the fundamental problem with the existing Arrow-Debreu model. In an Arrow-Debreu model, the risk-free rate is more fundamental than simply being the assumed funding rate for market participants in a risk-neutral world. No-arbitrage, or replication, arguments are used to establish risk-free portfolios that must earn the risk-free rate. Yet, since the theoretical basis for this rate is non-monetary Fisherian time preference, attempts to adjust the risk-free rate for potential funding problems are fundamentally at odds with the theoretical structure itself and need to be abandoned.

7.5 Transactions and Search Costs

In effect, neoclassical theory leaves ‘exchange arrangements indeterminate’ and is ‘forced to treat money metaphorically as potential consumer or producer good, which, in fact it is not’ (Niehans, 1978, p. 99). Taking a commodity view of money, and denying the existence of credit money, exchange arrangements can be added by considering all commodities ‘as potential media of exchange’ (Niehans, 1978, pp. 99-100). The medium of exchange solves the fundamental problem of barter: the double coincidence of wants (Jevons, 1875). Nevertheless, transactions costs have an impact on liquidity. ‘The ease of conversion of an asset into something else suggests the importance of transactions costs. These, in turn, may be formulated into terms of search theory’ (Hahn, 1990, p. 64). Such models attempt to explain the nature of money.

Neo-Walrasian general equilibrium models, in which all trades are conducted at the first point of time (time-0), do not have to contend with uncertain payment times; this form of liquidity risk is absent. The Walrasian auctioneer provides a central exchange that removes the need to search for a suitable counterparty (Vayanos & Wang, 2011, p. 292). Money, by contrast, provides utility for agents facing random payments times (Patinkin, 1965, p. 80). This technique can be questioned for its ‘failure to capture a large part of reality’ (Niehans, 1978, p. 18, n. 44), for, if necessary, agents can negotiate between themselves to eliminate payment uncertainty. Instead, Black (1970a, p. 7) observes that a ‘person or institution does not need to hold liquid assets to help handle unexpected cash flows. Almost all cash flows are expected, and the cost and risk associated with handling those cash flows that are unexpected are very low.’ Similarly, Niehans (1978, pp. 18-19) concludes that uncertainty concerning the timing of payments is neither a necessary nor sufficient ingredient for monetary theory. For Niehans, the medium of exchange is the commodity with the lowest transactions and storage costs, and its use provides efficiency gains (Niehans, 1978, p. 18). In this view, the social invention of money somehow reduces transactions costs.

Consider a neo-Walrasian economy with the explicit friction that agents can only meet for exchange randomly. Some of the necessary Arrow-Debreu contracts for a Pareto optimum cannot arise (Kiyotaki & Wright, 1989, p. 931). The potential for credit payment to solve the barter problem is disallowed, and the emphasis is placed on the ‘different intrinsic properties’ (Kiyotaki & Wright, 1989, p. 928) of each of the monetary commodity ‘candidates’, in accordance with the commodity view of money. ‘A commodity is regarded

as money for our purposes if and only if it can be traded directly for all other commodities in the economy' (Clower, 1967, p. 5). In equilibrium, agents converge on holding the commodity, for which they have no immediate need, but which has the lowest storage cost. The commodity with the lowest transactions and storage costs becomes a dominant media of exchange, hence marketable. Commodity 'money' spontaneously appears as the result of optimising agents' minimising their storage costs and allowing indirect trade (see Kiyotaki & Wright, 1989; Kiyotaki & Wright, 1993).

The commodity acting as medium of exchange, however, is implicitly in a raw, un-minted state, that is, not in any monetary form. It is nevertheless held for reasons unrelated to consumption or production (Kiyotaki & Wright, 1989, p. 937). Although trade could conceivably be conducted by credit or other forms of financial asset, these potential media of exchange are excluded by construction.[43] No recognisable form of money appears in these models. Instead, the agents play an optimising game of commodity middlemen. Other configurations of the model produce equilibria where there is more than one 'medium of exchange', because, of the three types of agent in the model, only one pair can trade directly, the other two pairs must use different commodities as a store of value for future trades (Kiyotaki & Wright, 1989, p. 938). One commodity is used as a medium of exchange even though it has the highest storage costs because of its 'superior marketability'. High storage costs, interpreted as negative returns, are used to give justification that, in the real world, money has a rate of return dominated by all other assets (Kiyotaki & Wright, 1989, p. 938). In many cases 'money' fails to arise at all (Kiyotaki & Wright, 1989, p. 939).

Fiat money, which is 'intrinsically useless' (Kiyotaki & Wright, 1989, p. 951), is introduced into the model seemingly by helicopter and becomes the '*general* medium of exchange' (Kiyotaki & Wright, 1989, p. 942, original emphasis). The initial introduction of this money is entirely glossed over (Kiyotaki & Wright, 1989, p. 942, n. 12); the fiat money is accepted simply because it is trusted. Since there is no facility, except time, by which trust can be established, the model assumes the very proposition that it attempts to prove. There is no explanatory content in this treatment of fiat money, nevertheless the 'tenuousness' of fiat

[43] With credit payments and Rational Expectations, the economy could still be Walrasian even when agents only meet at random. If each agent knew the probability of meeting, as they would with Rational Expectations, then Arrow-Debreu contracts could be constructed.

money in the model is declared to be a strength, since, under the commodity view, the value of fiat money must depend on faith (Kiyotaki & Wright, 1989, p. 943).

Search costs affect expected returns in such a way that assets can be ranked in terms of liquidity and assets assigned liquidity premia. Lippman and McCall (1986) produce such a measure of liquidity, which can be used to compare assets. Liquidity is defined as 'the optimal expected time to transform an asset into money' (Lippman & McCall, 1986, p. 44). In their model, the time it takes to be converted into money depends on the variability in sequential price offers 'drawn from the same known distribution' (Lippman & McCall, 1986, p. 46). In this search-based model, liquidity ultimately depends on the asset's price variability and 'liquidity is determined by characteristics of the asset; characteristics of the seller have virtually no impact' (Lippman & McCall, 1986, p. 47). The association of liquidity with marketability (Lippman & McCall, 1986, p. 48, n. 12) results in the counter-intuitive conclusion that rising interest rates or time preferences lead to increased liquidity (Lippman & McCall, 1986, p. 47). The explanation is that the seller's optimal sale price falls as the discounted future costs of holding the asset fall with the rise in interest rates. Given the same distribution of offers, the likelihood of an acceptable match increases.

Transactions and search theories also have a problem with realism: search difficulties rely on the existence of clearly profitable niches that, inexplicably, remain unoccupied:

> There is of course a connection here between the search-theoretic and straightforward transaction cost approach. Ordinary economic theory suggests that when there is inducement to search there will also arise middlemen who can profitably reduce search costs. Ordinarily agents then do not search but pay a middleman, which is the transaction cost of the theory. Middlemen, in turn, will charge more in markets where the frequency of offers and bids is low. As I have already noted, such a story is not only required by ordinary observation of the world but also by economic theory which supposes that profitable niches do not remain unoccupied. (Hahn, 1990, p. 73)

Search and transactions costs can be mitigated by dealers and market makers. As seen in Section 5.2, market makers provide inventory services for profit and, in doing so, take liquidity risk, but avoid price risk where possible. Each market maker can add market liquidity to any asset by relying on a correlated hedge asset used by all other market makers. Liquidity and bid-offer spreads are determined by the frequency of the asset's turnover,

search and transactions costs, or ‘return characteristics’ (Ho & Stoll, 1981, p. 62).

Walrasian price flexibility and market makers relate to the distinction between ‘flexprice’ and ‘fixprice’ markets (Hicks, 1974). Prices in flexprice markets are determined by supply and demand, whereas in fixprice markets prices are less responsive to supply and demand imbalances. The neo-Walrasian, perfect market liquidity setting aligns with the flexprice assumption. Notably, a flexprice market is characterised by intermediate traders (Hicks, 1974, p. 24) who ‘expect to make a profit’ from their activities (p. 26). The neo-Walrasian framework has, by its very nature, an implicit dealer structure underpinning it, although the ‘basic Walrasian story abstracts from the actual mechanics of markets’ since ‘buyers and sellers are all present at the same time’ (O’Hara, 2003, p. 1336). The Walrasian auctioneer can be interpreted as the single market maker that serves an ideal market (see Section 5.2).

To address the assumption of perfect market liquidity, many empirical studies have investigated the effect of transactions costs and other measures of illiquidity on asset prices. This literature can be distinguished from that providing the ‘micro-foundations’ for commodity money based on transactions and search costs (see Section 7.5). As with the equity premium and risk-free rate puzzles, general equilibrium models cannot explain the volume of trading observed in reality. In models with symmetric information, each agent’s best interest is served by holding the market portfolio, and not trading at all, a strategy that can only be reinforced by transactions costs. Nevertheless, the idea of transactions costs is used as the basis for analysis of the market ‘microstructure’ and the effects of illiquidity by studying measures constructed from trading volumes and daily price changes.

Fischer Black’s description of a liquid market (see Section 5.2) shows that large orders to buy or sell cannot be fulfilled without large price movements, which create ‘a positive relationship between the order flow or transaction volume and price change’ (Amihud, 2002, p. 33). One interpretation of market liquidity prompts illiquidity measures based on a ratio of absolute returns and turnover: ‘the daily price response associated with one dollar of trading volume’ (Amihud, 2002, p. 32). Illiquid stocks are then those with higher average absolute returns for any given dollar volume. A stock is illiquid ‘if the stock’s price moves a lot in response to little volume’ (Acharya & Pedersen, 2005, p. 386) and the cost of transacting illiquid stocks is higher because order flow has more of a price impact. In another interpretation, illiquidity could be signalled by unusually large price reversals following a day with a high volume of trading (Pastor & Stambaugh, 2003). A more

comprehensive measure, which attempts to include 'trading quantity, trading speed, trading cost, and price impact', defines illiquidity as 'the standardized turnover-adjusted number of zero daily trading volumes over the prior 12 months' (Liu, 2006, p. 631).

These empirical measures of market liquidity suffer conceptual problems, and each is 'a fundamentally different measure' (Cochrane, 2005a, p. 10). Observed correlations between asset-price changes and their order flow or traded volumes are strong, but measures of liquidity put forward in the literature often differ fundamentally or are contradictory, and these difficulties prevent the emergence of an agreed measure of market liquidity. As a result, evidence as to whether transactions costs affect expected returns is mixed (O'Hara, 2003, pp. 1338-1339), but many studies find that 'illiquid' stocks and bonds display an expected return premium attributed to illiquidity (Acharya & Pedersen, 2005; Amihud, 2002; Amihud et al., 2005; Liu, 2006; Pastor & Stambaugh, 2003). The empirical literature on market liquidity concludes that stocks with higher transactions costs—those deemed illiquid—should have a higher expected excess return as compensation for the costs of market illiquidity.

Illiquidity costs can be incorporated into the consumption-based asset-pricing model in several ways. In the simplest case, when investment horizons are known and transactions costs are constant, the asset's payoff is simply reduced by the sale cost and asset prices adjust to compensate for this lower expected payoff (Acharya & Pedersen, 2005). 'The liquidity premium on the risky asset with trading costs is then defined as the decrease in its expected return that would leave the investor indifferent between this asset and an identical asset with no trading costs' (Amihud et al., 2005, p. 292). When agents have variable time horizons, those with shorter horizons hold those assets with lower transactions costs, whereas agents with longer horizons can amortise costs over a greater period and are less concerned with minimising them (Amihud et al., 2005, p. 282).

Liquidity-adjusted extensions of the standard systematic pricing workhorse model, the CAPM, include asset-specific and overall market or aggregate illiquidity costs (Acharya & Pedersen, 2005). The introduction of market illiquidity provides the necessary systematic or aggregate factor to deduce that assets which 'are more sensitive to aggregate liquidity have substantially higher expected returns' (Pastor & Stambaugh, 2003, p. 683), and that liquidity is a risk factor and 'a good candidate for a state variable' in pricing (Pastor & Stambaugh, 2003, p. 643). In addition, the basic CAPM extended to include aggregate or

market liquidity captures the important differences in cross-sectional returns (Liu, 2006) and removes the need for the size and book-to-value factors in the Fama-French model (Fama & French, 1996). Expected returns in the liquidity-based asset-pricing model of Holmström and Tirole (2001) are based on the asset's sensitivity to 'aggregate liquidity', which is 'the aggregate value of financial instruments used to transport wealth across time and to back up promises of future payments' (Holmström & Tirole, 2001, p. 1837, n. 1). Examples of aggregate liquidity used to support the model are land or Treasury bonds backed by land. In each case, the payoff is exogenous. In this model, liquid assets take more of a collateral role, whereby land must be available to fund new projects.

Several conceptual problems appear with the aggregate transactions cost approach. First, the market portfolio underlying the CAPM is the 'wealth portfolio' or a claim on all future consumption. It includes all physical and human capital (Cochrane, 2005b, p. 169). The idea of relative market illiquidity must conceptually represent the ease of trading all wealth—a difficult concept to apply in practice and therefore of dubious realistic merit. In the empirical literature, aggregate liquidity is often an average across all the individual asset's liquidity measures, or the cost of a trade distributed evenly across all assets (Pastor & Stambaugh, 2003, p. 650). Second, asset-specific transactions costs simply reduce the asset payoff, and the predictions from the liquidity-adjusted CAPM are identical to the non-adjusted version (Acharya & Pedersen, 2005, pp. 380-381). Although the decomposition provides new points of traction for empirical testing, no explanatory power is added. Third, a premium for illiquidity is not the same as a premium for liquidity; the difference is looking for a price discount in an 'illiquid' security compared to a frictionless benchmark, versus the 'convenience yield' or price premium for money-like assets (Cochrane, 2005a, p. 7). Finally, although sitting in a general equilibrium model, the analysis is essentially microeconomic. Macroeconomic liquidity is represented by the aggregation of individual asset liquidity. Transactions costs, which in reality form the basis of dealer profits, are simply exogenous costs. As noted in Section 5.6, aggregate liquidity constructed from market liquidity suffers from a fallacy of composition. The deficiencies of this analytical approach are evident in the contradictory concerns about the excess liquidity resulting from QE and the absence of market liquidity displayed by the asset markets (PwC, 2015).

In principle, the observation that two assets with very similar payoffs can have different expected returns suggests an arbitrage opportunity. Once the 'bid-ask spread, brokerage

fees and cost of carrying the short position' are taken into account, however, the 'apparent arbitrage profit disappears' (Amihud et al., 2005, p. 333; see also Cochrane, 2005a, pp. 2-3). Cochrane (2005a) surveys the evidence that restrictions on short sales similarly affect the market's ability to maintain arbitrage relationships between financial options and their underlying assets. Transactions costs can affect asset payoffs and therefore asset prices. This finding should not be controversial, but also should not be overemphasised. Instead of transactions costs having their effect on the pricing kernel, they can be categorised as cash flows of the asset. What is less certain is whether transactions costs can explain cross-sectional returns or time-series variability (O'Hara, 2003, p. 1350). In models with transactions costs, 'securities with the same dividend streams have different prices if they have different transaction costs' and the existence of a common pricing kernel is called into question (Amihud et al., 2005, p. 278). For illiquidity to move from a being a cost to being a risk and thereby affecting the stochastic discount factor or pricing kernel, it needs to be time varying or systematic (O'Hara, 2003, p. 1339).

The existence of a liquidity premium arising from risk faces a fundamental challenge. 'In a world in equilibrium, it simply does not make sense to think of liquidity as an attribute of an asset that affects its expected return. There is no trade-off between liquidity and return' (Black, 1970a, p. 1). This assertion rests on two observations: first, that the risk of an asset can be entirely separated from its funding by financial derivatives; and, second, that 'liquidity, if it has any meaning at all, is associated with risk transfers rather than cash transfers' (Black, 1970a, p. 7). Interest rate, currency and credit default swaps allow the separation of an asset's various risk components from its ownership. Selling costs should be paid by the asset user, not the asset owner, and not affect the asset's return (Black, 1970a, p. 2). The risks associated with an asset can be entirely separated from its funding, and, in the limit, the investor simply provides a riskless loan with pre-arranged cash flows. As long as the collateral from the borrower is sufficient, the lender is indifferent to new information that affects the asset's price (Black, 1970a, p. 2). The significance of collateral, the interdependence between market and funding liquidity, and the effect of information on asset prices are discussed further below.

Ultimately, the illiquidity-premium literature identifies and describes the phenomenon that asset owners require compensation for expected selling costs, costs that permit the profits of market dealers. This illiquidity premium suffers from the problems outlined in Section

6.2, since it is built on implicit Wicksellian foundations, where asset returns are given by productivity, and allowing the comparison of two assets with the same pre-cost return. The illiquidity premium should be interpreted as the compensation for transactions costs that equilibrates expected returns across assets. It cannot be interpreted as a premium; the issuer of the financial asset does not pay the transactions costs.

The consumption-based asset-pricing model has the intuitively pleasing feature that payoffs in states where an extra degree of consumption would be welcomed are given a higher value weighting. This is the very crux of the risk-neutral 'probability' paradigm. Risk-neutral probabilities play a very important part in asset pricing. Although they are often referred to as probabilities they are better thought of as subjective value weightings or measures of the various outcomes associated with an asset payoff. Risk-neutral pricing can be understood with reference to the idea of certainty equivalence. Pratt (1964) considers the choice between an uncertain return R and its expected return $\mathbb{E}[R]$ with certainty. For a risk-averse investor with utility function $u(.)$ and wealth W to be indifferent between these two options requires a risk premium π such that

$$u\big(W(\mathbb{E}[R]-\pi)\big) = \mathbb{E}[u(WR)]. \qquad (7.1)$$

In other words, a risk-averse investor will value an investment with a certain return more highly than an uncertain investment with the same expected return. Risk-neutral probabilities are the result of applying this value weight to their equivalent real-world or physical probabilities.

Using the risk-neutral measure, each asset can be deconstructed into its component risks and each risk can be evaluated. Each asset can be assessed at an atomic risk component level and then compared to other assets, allowing for risk replication by other, related assets and the discipline of no arbitrage can be applied. This deconstruction provides the theoretical support for Black's (1970a) assertion that asset risk can be separated from funding risk, and that asset prices are based on risk transfers and not flows, stocks or liquidity.

The fundamental difficulty in connecting liquidity with funding stems from the irrelevance of liquidity and money to a neo-Walrasian equilibrium. Motivating transactions costs and trading in general equilibrium models requires: uninformed 'noise' traders (Black, 1986); that agents are unable to borrow (Amihud et al., 2005, pp. 283, 291); or other 'unpalatable

devices' (Cochrane, 2005a, p. 14). Such devices serve as 'as if' mechanisms to investigate the effects of illiquidity and trading in general equilibrium models, but do not purport to be deep explanations. Market liquidity is 'explained' by turnover, but the reasons for trading are not. Furthermore, the characteristic of market liquidity is viewed as an intrinsic feature of the asset itself and not due to the institutional arrangements of organised markets and dealers. The empirical studies showing that transactions costs affect asset pricing demonstrate that the balance-sheet capacity of market makers limits the absolute market liquidity of the neo-Walrasian model (Chordia et al., 2005). The effect on market makers is time varying, persistent and therefore provides predictive power.

Next, we see how the commodity view of money relates to the actions of the Walrasian auctioneer.

7.6 Cashless Models

In a frictionless, complete markets world all transactions are spot transactions and there is no delayed settlement and no problem with non-payment (Hicks, 1989, p. 47). Money is unnecessary as a medium of exchange and is always dominated as a store of value by an interest-bearing risk-free government bond. To bypass the problems of establishing the value of a fiat monetary asset, modern extensions of the classical approach restrict money only to the form of a unit of account, or *numéraire*. That general equilibrium models have no essential need for money can also be justified by the argument that, in reality, physical cash is becoming obsolete. The difficulty of adding money to general equilibrium models can be resolved by appealing to the idea that moneyless models represent the natural outcome of evolutions in institutions and technology:

> For years, economists have had difficulty in incorporating money into rigorous general equilibrium models. To the elegance of the Walrasian model of an exchange economy has been bolted on an assumption about the technology of making payments such as a "cash in advance" constraint. These untidy ways of introducing money into economic models are not robust to changes in institutions and technology. Is it possible that advances in technology will mean that the arbitrary assumptions necessary to introduce money into rigorous theoretical models will become redundant, and that the world may come to resemble a pure exchange economy? (King, 1999, p. 48)

These ‘cashless’ models can be supported by the observation that, in recent times, the use of physical cash has given way to credit cards and mobile-phone payment technologies, and therefore represent an inevitable technological ‘cashless limit’ towards which society is moving. It is claimed that the analysis of models without ‘monetary frictions’ provides insights into the hypothetical limiting case of this technological process (Galí, 2008; Woodford, 2003). Galí (2008, p. 24, n. 4) dismisses the *ad hoc* approach of including real money balances in utility functions, arguing that the cashless model is sufficient for the purposes of monetary analysis.

Nevertheless, in cashless models the representative agent can invest a portion of their wealth in balances of base money, which pay interest at the central bank’s policy rate. Here the line between neo-Walrasian and neo-Wicksellian models blurs. The agent only holds non-zero money balances if the policy rate equals the Wicksellian natural rate of interest (Woodford, 2003, p. 69). In addition, Rational Expectations ensure that the *ex-ante* version of the Fisher Equation determines expected inflation (Woodford, 2003, p. 50). Central bank monetary policy has temporarily non-neutral effects due to ‘Keynesian’ sticky prices which mean that ‘the real rate of return at which borrowing and lending occurs can differ from the natural rate of interest’ (Woodford, 2003, p. 51).

Although these models supposedly ‘resemble the channel systems’ used by central banks (Woodford, 2003, p. 75), the explicit payment of interest on base money is not the most remarkable aspect of a channel system. As we have seen in Section 4.6, the more interesting feature is that the central bank is not required to be an active participant in the money market. The channel system allows the central bank to enforce its policy rate in the interbank market, even though overnight balances at the central bank need not be significantly different from zero, simply by changing its borrowing and lending rates. By building the model on interest paid by the central bank at the Wicksellian rate, the cashless models assume what they set out to explain, but this difficulty is brushed aside because the assumption of interest paid on base money provides a convenient ‘as if’ means of establishing price determinacy in a cashless, neo-Walrasian world (Woodford, 2003, p. 83).

The cashless models, uninhibited by monetary frictions, are the apotheosis of a neo-Walrasian model: complete markets and perfect market liquidity (Mehrling, 2011, p. 64). In the monetary analysis of Galí (2008) and Woodford (2003), the standard model is ‘cashless’, there is no requirement for money except as a unit of account. The result is an

'accounting system of exchange' (Fama, 1980, p. 39) where exchanges are direct and measured only in the *numéraire*. 'Money [or unit] of account makes possible prices and debt contracts, which are all that are required for extensive multilateral exchange to take place' (Ingham, 2004b, p. 6). The *numéraire* need not have any physical presence in a general equilibrium model; the only function that the unit of account is permitted is the limited role of converting relative prices to 'absolute' or 'money' prices (Patinkin, 1965, p. 16). A neo-Walrasian general equilibrium can determine prices in any arbitrary *numéraire*, that is, an accounting price, but that these prices are not economically significant (Rogers, 2006b, p. 6), since only relative prices, or price ratios, are determined by the model. The unit of account 'is defined in terms of a claim to a certain quantity of a liability of the central bank, which may or may not have any physical presence' (Woodford, 2003, p. 63). The overall price level, which can be denominated any fictitious measure, is not just indeterminate, it is conceptually undefined.[44]

The ontology of money outlined in Chapter 3 certainly allows for a unit of account to exist independently of any money object, and the neo-Walrasian general equilibrium can be interpreted as a representation of this phenomenon. The neo-Walrasian general equilibrium, however, represents a framework in which all contracts are promises to pay in a specified consumption good; an interpretation inconsistent with the credit theory of money. The Walrasian auction can be interpreted as instantaneous credit clearing at the beginning of time, after which no financial contracts exist that promise to pay in an abstract unit of account (Rogers, 2006c). The *numéraire* is an arbitrary commodity that acts as a unit of account for the pricing of contracts, but once these prices are used in market clearing, it plays no role in specifying the nature of the promise made by each contract.

Finally, the view that society is becoming cashless is a superficial reading of the phenomenon of electronic payments, for behind each technology are money transfers in substance identical to that performed by cheques. For instance, Cochrane (2014, p. 199) thinks it perfectly feasible that 'you could buy a cup of coffee by swiping a card or tapping a cell phone, selling two dollars and fifty cents of an S&P 500 fund'. Instantaneous real-time and netted payment systems would effectively remove the need for reserves when 'crediting the coffee seller's two dollar and fifty cents mortgage-backed security fund.'

[44] See Rogers (2006c, 2008a) for a full discussion of the conceptual problems of cashless models.

Similarly, King (1999, p. 48) foresees that settlement could be conducted in liquid instruments 'for which there were market-clearing prices in real time'. In each view, it is anticipated that technology would enable perfect market liquidity, making all other forms of liquidity redundant. In addition, there would be no need for bank liabilities to act as money. Banks would lose their role as clearing houses of credit. Overlooked in these hypothetical worlds is the need for market makers and the elasticity of the real-time payments system to make this exchange of coffee without money possible. As Rogers (2008a) observes, it would be difficult for the economy to become truly moneyless in the neo-Walrasian sense, since 'e-money of some form to approximate the time-0 auction would require the instantaneous flow and coordination of information that is beyond the capability of even the most super of computers' (p. 11). A move to a neo-Walrasian moneyless payments system would involve the elimination of the central bank altogether, and the adoption of a free banking system (Rogers, 2008a, p. 24). Since each monetary space would still need a unit of account provided by an authority, the society could be cashless, but not moneyless (Ingham, 2004a, p. 179).

7.7 Summary

Although proxies for liquidity and money can be added to general equilibrium models, they provide no explanatory power. An intrinsically worthless fiat currency can be given value in the model only as a welfare-reducing exchange restriction, since a non-interest earning money asset is dominated by all other interest-earning assets. Consequently, 'monetary theorists are forced to proceed without sound theoretical foundations' (Rogers, 1989, p. xvii). Since money has no natural place and perfect market liquidity is assumed, any alleged concept of liquidity in the model bears no resemblance to the real-world concept of converting assets into money or even the pure liquid nature of money itself. The use of general equilibrium models to analyse liquidity is an example of the use of an ordering framework at the expense of the meaning of the concepts in a real-world sense (see Section 2.2). The neo-Walrasian general equilibrium model is inappropriate for analysing the spectrum of liquidity identified in Chapter 5. Search theories and pockets of inefficiencies are the only means by which 'liquidity' can be assessed.

The financial frictions literature sets out to explain the conditions that prevent the attainment of the general equilibrium as a concession to illiquidity. The assumption that a

stable and unique equilibrium exists brings with it the concept of a fair value of assets and the whole problem of 'liquidity' reduces to one of two problems. The first is of finding agents who need to trade at the same time, since, with symmetric information, there is no difficultly in establishing the correct price. The second concerns the transactions costs associated with selling. In this case, the fair value cannot be attained directly, but an intermediary or market maker can match buyers and sellers for a fee. As such, financial frictions are based on the difficulty of search and matching, not a problem of valuing the assets. In a general equilibrium model there is no doubt that a market-clearing price exists. Financial frictions slow the process of arriving at equilibrium, but do not alter the properties of the equilibrium.

Attempts to add money to neo-Walrasian models result in such severe distortions to the concepts of liquidity and money that any insights gained are of dubious value. Furthermore, the prevailing view that money is a commodity, even in its fiat form, thwarts the fundamental integration of the banking system. By adopting the commodity view of money, Real Analysis makes liquidity the spontaneous property of a particular commodity or asset, and thereby restricts the analysis of money as a liability. The only form of liquidity allowable into the framework is market liquidity, a characteristic of the commodity that becomes the medium of exchange. That market liquidity is provided by profit-seeking dealers is overlooked.

Modern asset-pricing theory, with its neo-Walrasian roots, operates in a world where the timing of payments is never an issue. The specification of Rational (or model-consistent) Expectations imposes the condition of ergodicity and collapses the dimensions of time and space. Since time is equivalent to place and all transactions happen at once, there cannot be the sort of credit-valuation problems associated with the credit view of money, whereby the value of credits payable is determined by the value of credits available to make payment. These concepts have no place in the models.

Furthermore, it is evident that the neoclassical 'cashless' benchmark model also suffers from many theoretical problems. Its reliance on Wicksell's natural rate of interest restricts the analysis to that of a single good (indeed, see Woodford, 2003, p. 64), and the reliance on the Fisher Equation for inflation determination is problematic. The emphasis given in Section 3.7 to the independence of unit of account function is in sympathy with the role of an abstract *numéraire* in neo-Walrasian general equilibrium theory, but with important

points of difference. That liquidity is not fundamentally derived from the unit of account will be discussed in Part IV. In neo-Walrasian theory, however, the unit of account is the only monetary function permitted, so it follows that liquidity is ejected along with the rest of the monetary functions.

In neo-Walrasian models, liquidity is interpreted as a constraint or incompleteness in the geographic landscape of the economic realm: some Pareto optimising transactions cannot be conducted, for some reason. Instead of being an issue that arises from commitments made now to make payments in an uncertain future, liquidity is understood to concern difficulties in making payments now. The possible interpretation of the Walrasian auctioneer as a representation of a perfect credit-clearing operation does align somewhat with the ontological view of the banking system presented in Chapter 4, but without the risk that the credit structure might malfunction—all credits are paid with surety—the value of any insights derived is undermined. Attempting to build a theory to explain the observed desire to hold liquid assets based on failed attempts to make current payments is not a promising avenue.

We must be mindful of these useful aspects as we now consider the relationship between liquidity and money in theories associated with Monetary Analysis. Post-Keynesian theories of money, banking and liquidity are good starting points for more considered treatments of liquidity, as set out in the next two chapters.

Part IV Monetary Analysis

Part IV uses the ontology established in Part II to consider liquidity and its relationship with money within the tradition of Monetary Analysis. Unlike Real Analysis, the role of money in Monetary Analysis is fundamental, the axiom of money neutrality is rejected, and the spectrum of liquidity identified in Chapter 5 is more readily visible. Yet pre- and post-GFC developments show that the treatment of liquidity and money is still problematic and areas of deficiency remain. The purpose of Part IV is to identify the elements of the theories in the tradition of Monetary Analysis that explain the phenomena of liquidity and money as defined in this study. Part V combines these elements with those discussed in other chapters into a conceptual framework.

Three branches of Monetary Analysis are assessed: the theory of liquidity preference in Chapter 8, and the theory of the monetary circuit and the theory of endogenous money in Chapter 9. Each can be broadly associated with the post-Keynesian school of economics, although they are in apparent conflict and attempts at reconciliation are not entirely successful (Chick & Dow, 2002). Suggestions for reconciliation with each other and elements of the theories from Real Analysis are made in Part V.

8 Liquidity Preference

8.1 Introduction

Keynes's theory of liquidity preference (Keynes, 1936) makes a clean break from the interest and asset-pricing theories based on pure exchange and general equilibrium. The theory of liquidity preference stipulates that the rate of interest is a monetary phenomenon, not one that is driven by the real forces of productivity and thrift as in the loanable funds theory. According to Chick (1983, p. 174, original emphasis), 'Keynes argued that his theory of interest follows naturally from what interest *is*: the price of having money now instead of later; the price of being liquid.' In opposition to the loanable funds theory of interest, Keynes's theory of liquidity preference states that interest rates are not a reward for saving, but instead are a reward for not hoarding money:[45]

> It should be obvious that the rate of interest cannot be a return to saving or waiting as such. For if a man hoards his savings in cash, he earns no interest, though he saves just as much as before. On the contrary, the mere definition of the rate of interest tells us in so many words that the rate of interest is the reward for parting with liquidity for a specified period. For the rate of interest is, in itself, nothing more than the inverse proportion between a sum of money and what can be obtained for parting with control over the money in exchange for a debt for a stated period of time. (Keynes, 1936, p. 167)

The effect of time preference is only partially reflected in the rate of interest. Keynes conducts Monetary Analysis and generalises

> …the classical psychological concept of time preference by splitting it into two elements; the marginal propensity to consume and the state of liquidity preference. It is the latter which plays a key role in determining the rate of interest in a monetary economy and it is this element which is missing from real analysis.

[45] See Robertson (1940, p. 16) for a loanable funds rebuttal that the rate of interest has an important role in balancing consumption and saving, and can instead be a reward both for saving and not hoarding. For Moggridge (1992, p. 596), however, loanable funds and liquidity preference are 'equivalent ways of describing the same phenomena.'

(Rogers, 2008b, pp. 10-11)

A reward for not hoarding money is required because of the 'convenience of liquidity' (Keynes, 1936, p. 168) which mitigates the 'existence of *uncertainty* as to the future of the rate of interest' (Keynes, 1936, p. 168, original emphasis). The risk in holding bonds, for example, is that 'if a need for liquid cash may conceivably arise before the expiry of *n* years [the length of the bond], there is a risk of a loss being incurred in purchasing a long-term debt and subsequently turning it into cash, as compared with holding cash' (Keynes, 1936, p. 169). If an asset with capital risk needs to be converted into money at some point in the future, its value at that time is uncertain. In essence, cash provides protection against asset-price risk.

According to Keynes, four motives drive the demand for liquidity: the transactions, finance, precautionary and speculative motives.[46] The transactions, finance and precautionary motives are assumed to be largely interest-inelastic; only the speculative motive is sensitive to the rate of interest (Keynes, 1936, p. 199). The transactions motive covers the liquidity required to conduct current exchanges and can therefore be interpreted as the opening demand in the monetary circuit (see below, Section 9.2), with its demand related to the cost of output or income. Similarly, the finance motive is the demand for money to begin an investment project, and, as such, is the investment equivalent of the transactions motive (Bibow, 1995), since it reflects the fact that planned investment requires finance to begin production (Keynes, 1937b, p. 246).

The precautionary motive represents a desire to hold a reserve of liquid assets as a protection against uncertainty. The precautionary motive provides 'for contingencies requiring sudden expenditure and for unforeseen opportunities of advantageous purchases, and also to hold an asset of which the value is fixed in terms of money to meet a subsequent liability fixed in terms of money' (Keynes, 1936, p. 196). It is here that an emphasis on liquidity over money must be recognised. As argued in Section 5.7, a liquid reserve need not consist solely of the medium of exchange itself; assets with a price-protection feature, convertible into the medium of exchange with little price risk, will serve the same purpose. The precautionary motive is satisfied not just by money-proper, but also by short-term

[46] For the transactions, precautionary and speculative motives, see Keynes (1936, p. 170); for the finance motive see Keynes (1937b).

assets that can provide 'portfolio diversification for precautionary purposes' (Chick, 1983, p. 202). It is a mistake to focus solely on money-proper, but, since many price-protected assets also yield a rate of interest, care must be taken in interpreting the precautionary motive and the determination of the rate of interest.

Finally, the speculative motive displays the most sensitivity to the rate of interest. The holdings of liquid assets that an investor demands depend on how far the current rate of interest differs from the investor's perceived 'safe' level of the rate of interest (Keynes, 1936, p. 201). Keynes first presents the theory of liquidity preference in a two-asset model comprising only two types of asset: 'money' or 'cash'; and 'debts' or 'bonds'.[47] Thereafter the theory is generalised to consider all assets. In the two-asset model, when the current rate is below the safe level, there is a strong desire to maintain holdings of cash, to avoid capital losses by delaying the purchase of debt until the rate has risen closer to the safe level. For those with the opposite belief, 'the purchase of debts will be a preferable alternative to holding cash' (Keynes, 1936, p. 170, n. 1). Therefore, the

> …individual, who believes that future rates of interest will be above the rates assumed by the market, has a reason for keeping actual liquid cash, whilst the individual who differs from the market in the other direction will have a motive for borrowing money for short periods in order to purchase debts of longer term. (Keynes, 1936, p. 170)

Importantly, the speculative motive depends on an organised market for trading assets, for 'in the absence of an organised market, liquidity-preference due the precautionary-motive would be greatly increased; whereas the existence of an organised market gives an opportunity for wide fluctuations in liquidity-preference due to the speculative-motive' (Keynes, 1936, pp. 170-171). The existence of organised markets, which can be interpreted as the presence of dealers providing market liquidity, presents an individual with the 'opportunity to revise his judgement and change investment' (Keynes, 1936, p. 153). As shown in Section 5.2, an organised exchange providing market liquidity gives investors some assurance of saleability, but Keynes highlights the uncertainty regarding asset prices at the time of sale. Organised markets allow for the existence of an observable interest rate

[47] Keynes uses the terms 'cash' and 'money' interchangeably, as he does with the terms 'debts' and 'bonds'.

that the public can compare to their subjective safe interest rates.

8.2 Liquidity Premium

In Keynes's generalised model, the total return expected from holding an asset is given by $a + q - c + l$, where a is the expected price appreciation of the asset measured in the unit of account, q is the yield or output of the asset, c is the carrying cost, and l is the liquidity premium (Keynes, 1936, pp. 225-228). The attributes q, c and l are measured in terms of the asset itself, so a is an adjustment factor necessary to convert the returns into units of account, and hence become comparable across assets.

The meaning of the liquidity premium l, however, is somewhat ambiguous and requires careful interpretation. Keynes describes it as the return that an investor is willing to forgo for the 'potential convenience' of 'the power of disposal over an asset' (Keynes, 1936, p. 226). On the surface, it is tempting to interpret this liquidity premium as a reference to the convenience of saleable assets with deep market liquidity, except that Keynes's premium is one of convenience given to liquid assets, not an extra return required to compensate for illiquidity. Keynes's liquidity premium is highest for money itself, whereas the premia for machinery, wheat and houses are likely to be 'negligible' (Keynes, 1936, pp. 226-227). The size of the liquidity premium for non-money assets is ambiguous, and so clarification must be made as to how Keynes's liquidity premium relates to market liquidity and the taxonomy outlined in Chapter 5.

Keynes uses the term 'liquidity' in both *A Treatise on Money* (Keynes, 1930), and the *General Theory* (Keynes, 1936) to put forward the theory of liquidity preference. As highlighted by Beggs (2015), subsequent interpretations of these usages have developed along two 'careers' or 'strands', and have consequently affected the understanding of liquidity preference. The *Treatise* strand found its home in the 'practical world of bankers, central bankers, and policymakers' (Beggs, 2015, p. 1), whereas the strand stemming from the *General Theory* was restricted to the realm of theoretical economics.

The second strand is considered in Section 8.3. The first strand, as stated in the *Treatise*, pertains in particular to the portfolio choices of bankers:

> Broadly there are three categories to choose from—(i.) Bills of Exchange and Call Loans to the Money Market, (ii.) Investments, (iii.) Advances to Customers. As a

> rule, advances to customers are more profitable than investments, and investments are more profitable than bills and call loans; but this order is not invariable. On the other hand, bills and call loans are more "liquid" than investments, *i.e*. more certainly realisable at short notice without loss, and investments are more "liquid" than advances. (Keynes, 1930b, p. 67)

Hicks uses 'more certainly realisable at short notice without loss' as a definition of liquidity that could 'be reinterpreted so as to give it a wider reference' (Hicks, 1989, p. 64). Again, it would be tempting to align this 'definition' with market liquidity and the ability to trade immediately with certainty in efficient markets. The difficulty with this characterisation is the interpretation of the final characteristic: without loss. The unresolved question is: without loss relative to what? For Hicks (1962, p. 790), the answer is that, for a liquid asset, the 'price at which it is realisable at short notice is much the same as that at which it is realisable at longer notice.' Hicks concludes that the term 'liquidity' refers to marketability (that is, market liquidity) relative to an absolute fundamental or long-period price, supported by an organised market of dealers (Hicks, 1962, p. 791).

The attempt to interpret liquidity preference within a framework built on market liquidity is a direct consequence of the commodity view of money outlined and rejected in Chapter 3. The view that money is simply the most saleable commodity limits the analysis of liquidity to market liquidity only. This ontological starting point leads Bronfenbrenner (1945) and Hicks (1962) to conclude that liquidity preference can be satisfied by a selection of assets from across a spectrum of assets ranked by market liquidity. To illustrate, consider Bronfenbrenner's description of 'perfect liquidity':

> By calling a commodity "perfectly liquid" in a given use we mean that it is *absolutely certain* that its holder can obtain for a unit of the commodity at least *one hundred per cent* of its (deflated) money cost to him minus only depreciation compensated for by actual use, measured in terms of general purchasing power, *immediately* upon deciding to dispose of it. (Bronfenbrenner, 1945, p. 407, original emphasis)

When money is merely the most traded commodity then market liquidity or saleability is the only form of liquidity that can be considered, since there is no other form of liquidity. Bronfenbrenner's interpretation of 'without loss' protects the holder from losses in value compared to the original price in real or price-deflated terms, but leaves unanswered what

the holder can obtain in return for the unit of commodity.

Hicks's long-period idea, on the other hand, introduces the unstated assumption of a fundamental value derived from a long-period or general equilibrium that can be used as a benchmark for the observed market price, with perfect market liquidity providing the frictionless benchmark. As seen in Section 7.5, neo-Walrasian general equilibrium theory is inappropriate for a deep analysis of liquidity; the fundamental value that Hicks relies on is undefined if money and liquidity are non-neutral and no long-period equilibrium exists. In practice, market liquidity is based on the existence of dealers who are prepared to provide continuously tradable prices for small amounts with a view to relative value. Dealers do not, as a rule, attempt to measure fundamental value (see Section 5.2). In the limit, when all assets have perfect market liquidity, money has no need to exist. Market liquidity cannot form the theoretical basis of liquidity preference.

In its reliance on dealer support, market liquidity is a form of liquidity that disappears in a crisis and is unable to satisfy liquidity preference at the very moment it is needed. Market liquidity is a mirage:

> This is the inevitable result of investment markets organised with a view to so-called "liquidity". Of the maxims of orthodox finance none, surely, is more anti-social than the fetish of liquidity, the doctrine that it is a positive virtue on the part of investment institutions to concentrate their resources upon the holding of "liquid" securities. It forgets that there is no such thing as liquidity of investment for the community as a whole. (Keynes, 1936, p. 155)

Market liquidity suffers from a fallacy of composition. It should be clear from the context that 'liquidity of investment' refers to market liquidity, as opposed to money-liquid assets, since '[i]nvestments that are "fixed" for the community are thus made "liquid" for the individual' (Keynes, 1936, p. 153). Organised markets, supported by dealers, allow long-term investments to be transformed into short-term investments for individuals, but not for the community as a whole. Market liquidity itself is not an 'intensive property' (Lozano, 2015, p. 16), instead it is a 'property described by functional relationship between delay time and percentage realization, and not by a single number' (Tobin & Golub, 1998, p. 14). This multidimensional and functional nature means that 'there may be no simple ordering of assets according to [market] liquidity' (Tobin & Golub, 1998, p. 14), and any ordinal relationship is necessarily complex.

Furthermore, the cost of market liquidity in organised markets does not justify a liquidity premium beyond that which covers the expected cost of trading. In liquidity preference, the 'underlying idea of the liquidity premium is that assets will have an "illiquidity discount" relative to money if their future price is less predictable than that of money' (Bibow, 1998, p. 245, n. 9). Crucially, the uncertainty relates to the fact that the 'price may go *either way*' (Bibow, 1998, p. 245, n. 9, original emphasis). Transactions costs and market illiquidity belong in the carrying cost component c, since it is 'a clear cut *one-way* guess' (Bibow, 1998, p. 245, n. 9, original emphasis). The measure of an asset's market liquidity—its bid-ask spread—would be seen in its carrying cost, not its liquidity premium.

In the absence of organised markets, and hence market liquidity, the speculative motive would not operate, although liquidity preference would still be driven by the other motives. Keynes distinguishes market liquidity from the form of liquidity embedded in money: 'If individual purchases of investments were rendered illiquid, this might seriously impede investment, so long as *alternative ways* in which to hold his savings are available' (Keynes, 1936, p. 160, original emphasis). These alternative ways involve 'hoarding or lending *money*' (Keynes, 1936, p. 160, original emphasis), and would be preferred without the existence of organised markets in which investment assets 'can be easily realised for money' (Keynes, 1936, p. 161). Therefore, although liquidity preference cannot be built on a preference for assets that have market liquidity, its operation is affected by the presence of organised exchanges that reduce the sale costs of assets. Liquidity preference explains the desire for the aggregate or macroeconomic liquidity available in assets displaying price-protection, thus justifying the sharp distinction between money and non-money assets in the two-asset model that Keynes first uses to present the theory of liquidity preference.

Instead of favouring assets that can be sold with limited 'marketing time and effort' or concerns about 'the depth of an asset's market' (Beggs, 2015, p. 15), it is market risk and market liquidity risk that create a preference for liquidity. Investors wish to hold assets with a price-protection feature because they are uncertain about future price variability. Liquidity preference describes the choice between near-money assets that mature before the investor's horizon and longer-term assets with capital-uncertainty that rely on market liquidity for conversion into the medium of exchange. Market liquidity does not clarify the concept of liquidity in the theory of liquidity preference, because a high degree of market liquidity does not indicate that an asset has a low degree of uncertainty regarding its future

price or return—the very reason given for liquidity preference. Exchange-traded stocks and shares generally have a high degree of market liquidity but also a high degree of price variation.

Liquidity preference also mitigates the uncertainty relating to the interest rate payable for 'selling' a newly created debt 'asset' to a lender. In other words, liquidity preference would also be a defence against problems with funding liquidity and the future cost of borrowing. It is applicable to the liability side of an investor's balance sheet as well as the asset side (Beggs, 2015, p. 24). A decision to maintain a stock of price-protected assets because of uncertainty regarding the cost of future borrowing is consistent with liquidity preference.

Thus, it is a mistake to align Keynes's liquidity premium with the illiquidity premium derived from the transactions and search costs theory of liquidity. It follows that Keynes's liquidity premium is not an increase in expected returns for illiquid assets as it is in the empirical literature on liquidity (see Section 7.5). Instead, the abstract liquidity premium associated with money sets the standard that all other assets must match, meaning that liquidity preference is a theory of the rate of interest, and not a theory of liquidity premia. The interpretation that this study adopts is that liquidity preference is a desire to hold assets that have price stability in terms of the unit of account, and therefore provide price-protection. Price-protection is the essential feature of liquidity preference due to the 'convenience of holding assets in the same standard as that in which future liabilities may fall due and in a standard in terms of which the future cost of living is expected to be relatively stable' (Keynes, 1936, pp. 236-237).

8.3 Portfolio Rebalancing and the Demand for Money

The second strand of liquidity identified by Beggs (2015) concerns the demand and supply of money itself. In this strand, the interpretations of liquidity preference theory explain the demand for money due to risk aversion (Tobin, 1958) and the determination of asset prices by portfolio rebalancing (Hicks, 1989; Tobin, 1969). The portfolio-rebalancing paradigm involves equilibrating the supply and demand of available assets, including money, where expected rates of return are given by each asset's marginal rate of productivity, so that 'the price of assets moved to ensure that all assets were held by someone' (Mehrling, 2011, p. 62). On the one hand, the portfolio rebalancing approach is Walrasian because all assets 'are assumed to be saleable at prices determined by the balance of supply and demand'

(Mehrling, 2011, p. 64). The implication for liquidity is that the 'demand for money is not a demand for the ultimate *liquid* asset but only a demand for the ultimate *riskless* asset, as all assets are assumed to be liquid' (Mehrling, 2011, p. 64, original emphasis). On the other hand, the approach is also Wicksellian due to its reliance on capital in value, not physical, terms and returns determined by the marginal productivity of capital (Rogers, 1989, pp. 117-124).

The assertion that portfolio rebalancing is Walrasian, just because all assets are tradable, is not entirely accurate. Walrasian analysis cannot specify asset returns without the model being over-determined, some prices cannot be 'traded' because asset returns are given by the Wicksellian set up of the model (Rogers, 1989, p. 53). These aspects mean that the expected return and variance of each asset is known. This prior knowledge of all asset returns ensures consistent expectations, thereby removing uncertainty and the need for liquidity. The portfolio rebalancing interpretation of liquidity preference suffers from the problems associated with Wicksellian and Walrasian theories identified in Chapter 6.

The most significant obstacle for portfolio rebalancing to be a valid interpretation of liquidity preference is that the speculative motive is not used to determine interest rates (Chick, 1991). Instead, it asserts that 'liquidity must be regarded as an explanation of the existence and level not of the interest rate but of the differential between the yield on money and the yields on other assets' (Tobin, 1958, p. 65, n. 2). The portfolio rebalancing theory loses the key element of the liquidity preference theory—that uncertainty about future rates determines current interest rates and a demand for liquidity (Chick, 1983, pp. 213-214). The speculative motive is unnecessary, since returns are already determined, all that is left is the precautionary motive as a utility-derived selection based on asset volatility. The portfolio rebalancing theory is, at best, complementary to liquidity preference (Bibow, 1998, p. 260; Chick, 1983, p. 213).

Although liquidity preference has been interpreted as a demand for money, Keynes's definitions of liquidity and money are flexible and depend on investment horizons:

> …we can draw the line between "money" and "debts" at whatever point is most convenient for handling a particular problem. For example, we can treat as *money* any command over general purchasing power which the owner has not parted with for a period in excess of three months, and as *debt* what cannot be recovered for a longer period than this; or we can substitute for "three months" one month or three

> days or three hours or any other period; or we can exclude from money whatever is not legal tender on the spot. It is often convenient in practice to include in money time-deposits with banks and, occasionally, even such instruments as (*e.g.*) treasury bills. As a rule, I shall, as in my *Treatise on Money*, assume that money is coextensive with bank deposits. (Keynes, 1936, p. 167, n. 1, original emphasis)

Depending on the investor's time horizon, the definition of 'money' is flexible and the distinction between liquidity and illiquidity is similarly relative. With this relative distinction, Keynes is able to present the theory of liquidity preference using the two-asset model consisting of 'money' and 'debts', to represent liquid and illiquid assets, respectively. According to the ontology of money established in this study, debts of three months, for example, would not conform to the characterisation of money as a medium of exchange, since money is a debt-repayment token, and must be undated. Liquidity preference does not concern money *per se*; the difference between liquid and illiquid assets matters:

> The current rate of interest depends, as we have seen, not on the strength of the desire to hold wealth, but on the strengths of the desire to hold it in liquid and illiquid forms respectively, coupled with the amount of the supply of wealth in the one form relatively to the supply of it in the other. (Keynes, 1936, p 213)

The conclusion to draw is that Keynes uses the terms 'money' and 'liquidity' as synonyms in the context of liquidity preference. Liquidity preference is satisfied by holding assets 'in the form of immediate, liquid command', that is, 'in money or its equivalent' (Keynes, 1936, p. 166). The theory of liquidity preference rests on liquidity in general and not on money in particular; it is a theory of liquidity and not a theory of the demand for money. Liquidity preference can form the basis of 'a theory of financial asset pricing' (Brown, 2003, p. 330) or even a 'generalisation of the classical (marginal) theory of value' (Townshend, 1937, p. 160).[48]

Liquidity preference expresses a tendency to avoid assets with capital uncertainty in favour of those with a price-protection feature, but not just money-proper. Keynes's liquidity

[48] For attempts to apply liquidity preference to asset pricing see also Boulding (1944), Mott (1985), Townshend (1937), Wells (1983) and Wray (1991a, 1991b).

premium only applies to such liquid assets as short-dated, self-liquidating, near-money securities with a price-protection feature, and, in the modern context, this set of liquid instruments would include repo. Although this spectrum of liquidity in assets is juxtaposed with the sharp bifurcation of liquid and illiquid assets in Keynes's two-asset model, we can reject the assertion that, by adopting the two-asset model, 'Keynes invoked a simplifying assumption, namely, that bills, bonds, equities, or other securities can be treated for analytical purposes as holding the same relationship to money' (Brown, 2003, p. 330; see also Tobin, 1958).

Nor is liquidity preference a desire for less volatile instruments due to risk aversion, as it is in portfolio rebalancing. It is concerned with the uncertainty of the realised asset return if conversion to money is required unexpectedly, and so refers to a desire to hold price-protected or self-liquidation assets, instead of assets with market liquidity that can only be converted into money by means of market makers and organised markets. The key issue that portfolio-rebalancing theories fail to cover is the very risk that liquidity preference mitigates: that the market price of an asset falls at the precise moment that it must be converted to a means of payment. It does not cover the possibility that the investment horizon may change unexpectedly, and, since the asset must be liquidated before the planned investment horizon, its expected return cannot be realised. The fixed-term analysis of Hicks and Tobin fails to address this form of liquidity risk.

Hicks (1974, p. 37), therefore, realises that 'we need something more than a portfolio selection theory; we need a theory of liquidity.' For Hicks, an important addition to the theory would consider 'things which are unknown now, but will become known in time' (Hicks, 1974, p. 39). Portfolio construction is not a single choice activity, but involves decisions sequenced in time where optimal choices change with the arrival of new information, and, for Hicks, liquidity relies on time and the postponing of investment decisions: a waiting theory of liquidity (Bibow, 1998). It is rational that a store of liquid assets 'permits the individual to take advantage of currently unforeseeable, future opportunities, while simultaneously self-insuring the holder against untoward events' (Davidson, 1988, p. 335).

This sequence or waiting theory of liquidity can be interpreted within the asset-pricing framework of Equation (6.9) by means of 'managed portfolios' of the form $z_t\, x_{t+1}$, where z_t is a trading strategy that improves expected portfolio outcomes $\mathbb{E}[m_{t+1} z_t\, x_{t+1} | \mathcal{F}_t]$ when

compared to a simply buy-and-hold strategy (Cochrane, 2005b, pp. 133-134).[49] Managed portfolios allow the asset-pricing framework to handle the fact that asset prices do not follow a random walk and that discount factors change over time with the arrival of new information (Cochrane, 2005b, p. 131) by means of the law of iterated expectations:

$$\mathbb{E}[\mathbb{E}[x|\mathcal{F}_t]|\mathcal{F}_{t-1}] = \mathbb{E}[x|\mathcal{F}_{t-1}], \tag{8.1}$$

which expresses the idea that 'your best forecast today of your best forecast tomorrow is the same as your best forecast today' (Cochrane, 2005b, p. 133). In terms of holding money to delay an asset purchase, the expected and subjective return on holding money until time τ and then purchasing some asset with payoff x_T, with $\tau < T$, given what is known now, relative to what is expected to be known at τ, must be greater than simply purchasing x_T now.

Placing the waiting theory into the context of modern portfolio theory shows that it relies on marketable assets that allow for portfolio reallocation, and therefore has Walrasian roots, since it could easily be handled by a suitable number of Arrow-Debreu contracts or continuous trading (Lucas, 1984). Market liquidity available for portfolio re-allocation allows agents to respond to new information (Hicks, 1974). By emphasising the managed portfolio benefits of liquidity, the waiting theory relies on shiftability and the framework reverts to one based only on market liquidity, since investors prefer assets with marketability (Hicks, 1974, p. 42), or low transactions costs, so that they can perform subsequent changes in asset choice at minimum cost. Liquidity preference then becomes a preference for marketable assets over non-marketable assets, such as 'a new factory' (Hicks, 1974, p. 42), but Bibow (1998) rightly dismisses the choice between a new factory and an existing marketable security as an investment, not a portfolio, decision.

Hicks (1974, p. 43, original emphasis) recognises that an asset can be 'imperfectly liquid' if, at a random time of sale its price can be '*abnormally* low'. The risk of loss at sale is a factor in the ability to make sequential decisions. The key factor is not market liquidity but the possibility of loss at the time of sale; we need to consider the relationship between liquidity preference and the riskiness of assets.

[49] Technically z_t is previsible, that is, it is known or chosen at time t.

8.4 Theory of Bearishness

If liquidity preference is simply a response to concerns that asset prices will fall, then liquidity preference could be interpreted as an extension of Keynes's (1930) theory of bearishness (Brown, 2003; Ertürk, 2006; Skidelsky, 1992, pp. 561-562). As Keynes puts it, in an organised market, the 'price will be fixed at the point at which the sales of the "bears" and the purchases of the "bulls" are balanced' (Keynes, 1936, p. 170; see also Keynes, 1930a, p. 250). For a bear, the alternative option is to hold money, and 'each increase in the quantity of money must raise the price of bonds sufficiently to exceed the expectations of some "bull" and so influence him to sell his bond for cash and join the "bear" brigade' (Keynes, 1936, p. 171). The speculative motive, as well as being tied to the concept of a safe level of interest rate, expresses 'the object of securing profit from knowing better than the market what the future will bring forth' (Keynes, 1936, p. 170). Following this line, Chick (1983, p. 204) has speculators, with short-term investment horizons, as the instigators of the speculative motive, and downplays the idea of the safe rate of interest, instead emphasising a 'normal' rate to which the speculator expects the current market rate to return, with deviations from this normal rate being a source of profit.

An analysis of market liquidity, and its reliance on market makers, should make us wary of emphasising the role of speculators. The necessary condition for the speculative motive is an organised market, which, in turn, requires the support of market makers. It is a mistake that an 'economist might refer to a market maker as a "speculator"' (Black, 1971, p. 29). An overemphasis on speculators also introduces the misleading concept of fundamental value that conflicts with the reality of limited information and uncertainty (see Section 7.5). According to Friedman (1953b, p. 175), speculators add stability to asset prices by providing the countervailing forces that restore prices to their fundamental values when affected by speculative bubbles or over-pessimistic market crashes. As seen in Section 5.2, price moves on an organised exchange should be unpredictable in the short term, in other words, efficient. Speculators may well add stability, but, if so, they cannot be market makers.

Even within the two-asset model, several difficulties arise from the interpretation of liquidity preference as a theory of bearishness. First, short selling—selling bonds that are not currently owned now, in anticipation of buying them back more cheaply in the future—is a superior strategy for an individual with a belief that asset prices will fall. A bear trader

could be a short-seller and not just one 'who prefers at the moment to avoid securities' and hold bank deposits (Keynes, 1930a, p. 250). By conducting a short sale, the bear does not take a position in money-proper; the funds received from the sale of the bonds in the present are used to borrow the very bonds needed to complete the sale. The position is the exact counterpart of the bullish investor, who can borrow short term to purchase longer-term debts. Both bulls and bears can express their views by leveraged positions in the bond asset, without any cash position, just as per the CAPM (see Section 6.6).

Chick (1983, p. 198) recognises this problem: 'The speculative demand is made to look ridiculous because it seems to imply that speculators occasionally hold a totally barren asset when obvious alternatives are available.' For Chick, however, the problem is more to do with the existence of interest-bearing, short-term, liquid assets that are just as good as money-proper to mitigate problems with the timing of sales of longer-term assets. As already concluded, however, liquidity preference should be defined in terms of near-money assets. The issue of short selling is more problematic.

Ultimately, the interpretation of liquidity preference as a pure theory of bearishness can be rejected by noting that the return on short selling can also go either way—the same problem identified with building liquidity preference solely on market liquidity. A convinced bearish investor can take a short position but would have the same exposure to uncertainty as a bullish investor with complete conviction. Although it is not always practically possible to engage in short selling, a theory of liquidity based on bearishness is undermined by the theoretical existence of instruments that would provide ready substitutes. That liquidity preference cannot simply be a theory of bearishness is reinforced when it is realised that Keynes's theory of bearishness was constructed within a Wicksellian loanable funds model, which the theory of liquidity preference was explicitly developed to oppose (Smithin, 2006).

Instead, the element of uncertainty, and liquidity's ability to provide safety from uncertainty, must be emphasised (Cardim de Carvalho, 2010). Market liquidity underpins the speculative motive by permitting, not general equilibrium and fundamental prices, but a focus on the near-term. Organised 'liquid' markets create the possibility of short-term speculative trading strategies and not an evaluation of long-term investment prospects (Keynes, 1936, pp. 158-159). The investor's decision is not based on 'the *absolute* level of *r* [the rate of interest] but the degree of its divergence from what is considered a fairly *safe*

level of *r*, having regard to those calculations of probability which are being relied on' (Keynes, 1936, p. 201, original emphasis). Thus, the likelihood and size of the anticipated gain from purchasing an asset must outweigh the likelihood and size of loss expected at the time of sale, which itself is uncertain. The 'actuarial profit or mathematical expectation of gain calculated in accordance with the existing probabilities—if it can be so calculated, which is doubtful—must be sufficient to compensate for the risk of disappointment' (Keynes, 1936, p. 169). It follows then that, for any given set of expectations, there must exist an interest rate at which the investor's subjective expectation of gain and loss are in balance. This balance must also be in terms of risk-neutral probabilities to incorporate not just monetary loss, but 'disappointment' (see Section 6.6).

For Kahn (1954), however, the distinction between the precautionary motive and the speculative motive is less about organised markets and more about the investor's degree of conviction in his expectations. The speculative motive is more prevalent if there is 'complete conviction' in expectations, and the precautionary motive enters when expectations are subject to uncertainty, represented by a perceived unreliability of the investor's subjective probability distribution. The precautionary motive operates in conjunction with the speculative motive, when expectations about future interest rates are not held with complete conviction and gives the speculative motive 'something to bite on' (Kahn, 1954, p. 246). With complete conviction, a bull can do no more than invest entirely in non-money securities, regardless of interest rate (Tobin, 1958), yet if there is any uncertainty, a bull investor will hold some money as a precautionary measure. Although the speculative motive appears to be a pure theory of bearishness, it is the addition of uncertainty, and the associated precautionary motive, that introduces the motivation for investors to hold liquid assets (Keynes, 1937d).

In the context of a theory of the rate of interest based on uncertainty, the idea that any investor can know better than the market, and believe as much, can only be interpreted as signifying different degrees of opinion as to the future of the market. For example, Runde (1994, p. 134, original emphasis) views the relationship between the speculative motive and uncertainty as one where 'speculators do not choose particular levels of speculative balances *because* they are uncertain about future price movements', but instead 'because they regard future price movements in one direction as more probable than in the other.' In other words, the compensating price associated with the safe rate of interest will differ for

each individual, and none will necessarily agree with the 'market' price visible in an organised market. Agents will compare their compensating price for an asset with the market price to decide whether to purchase the asset or not. Deviations from each investor's safe rate inform the individual's decision on the amount of price-protected liquidity to hold in their portfolios.

Risk alone cannot explain liquidity preference. The balance of bulls and bears is in terms of risk, not liquidity, and liquidity is not conceptually the same as risk:

> I am rather inclined to associate risk premium with probability strictly speaking, and liquidity premium with what in my *Treatise on Probability* I called 'weight'. An essential distinction is that a risk premium is expected to be rewarded on the average by an increased return at the end of the period. A liquidity premium, on the other hand, is not event expected to be so rewarded. It is a payment, not for the expectation of increased tangible income at the end of the period, but for an increased sense of comfort and confidence during the period. (Keynes, 1979, pp. 293-294, original emphasis)

In the *Treatise on Probability*, Keynes (1921) introduces a truth relationship between proposition a and hypothesis h with probability p as $a/h = p$. For Keynes, however, new evidence h_1, even if it leaves the probability p unchanged, 'increases the *weight* of the argument' (Keynes, 1921, p. 71, original emphasis), meaning that the proposition is not more or less likely, but that the probability estimate is more reliable. 'New evidence will sometimes decrease the probability of an argument, but it will always increase its "weight"' (Keynes, 1921, p. 71). The new evidence must be 'relevant' in the sense that any part of the evidence can affect the probability, even if the new evidence as a whole and on balance does not (Keynes, 1921, pp. 55, 71-72). With this new evidence the weight of the argument $V(a/h)$ is increased to $V(a/hh_1)$.[50]

Crucially, probability and weight are independent, since 'the weighing of the *amount* of evidence is quite a separate process from the *balancing* of the evidence for and against' (Keynes, 1921, p. 74, original emphasis). Runde (1990, pp. 282-283), however, questions

[50] Weatherson (2002) shows how Keynes's intuitive concept of weight can be re-interpreted to be reflected in the theory of imprecise probabilities, whereby weight can be represented by an interval of probabilities. The wider the interval, the more uncertain the event is, and the less is its weight.

the conclusion that new evidence must always increase the weight of a probability with the counter argument that new evidence may alert us to more alternatives or possibilities, and thereby undermine our previous confidence. Uncertainty can arise 'when new evidence is acquired to the effect that there are more alternatives or a larger field of possibility than previously imagined' (Runde, 1990, p. 283). This counterexample relies on epistemological difficulties in determining the range of possible outcomes—a quite separate matter from assigning probabilities to them. It may not necessarily apply specifically to variations in asset prices. Nevertheless, Keynes's 'two-tier theory of belief' of probability and weight (Runde, 1994, p. 133) is needed to understand liquidity preference.

Finally, the rejection of liquidity preference's association with both market liquidity and risk premia, provides further support for the rejection of the commodity view of money. If a commodity were to gain such a liquidity premium, arbitrageurs could sell the commodity short and profit from the liquidity premium. This arbitrage is not possible when money is credit and issued only by recognised money issuers, such as the government, central bank or commercial banking system. In other words, price-protected, information-insensitive assets cannot be produced by just anyone, and liquidity cannot be arbitraged by risk decomposition to remove the trade-off between liquidity and return (see Section 7.5).

Keynes's liquidity premium differs from a risk premium in that it relates to a desire for liquidity to mitigate a lack of confidence, with 'the difference corresponding to the difference between the best estimates we can make of probabilities and the confidence with which we make them' (Keynes, 1936, p. 240). The liquidity premium in liquidity preference, stems from uncertainty over the rate of interest in the future, and is subject to Keynes's concept of the weight of evidence supporting a proposition. The liquidity premium l is a payment or acceptable reduction in yield for otherwise identical assets 'of equal initial value' (Keynes, 1936, p. 226) and provides a sense of comfort or utility. Time and uncertainty are necessary for money to have a fundamental place. Asset prices are, in substantial part, determined by uncertainty and economic agents' desire to maintain a store of wealth in a liquid form that can provide a level of insurance against this uncertainty.

8.5 Consistent Expectations and Equilibrium

Keynes states that, 'in equilibrium', the total expected return for each asset (including the liquidity premium) 'will be *equal*' (Keynes, 1936, pp. 227-228, original emphasis). For

example, Kahn (1954, p. 230) explains that agents will remain indifferent between short-term bills and long-term bonds if the expected fall in the bond price, or rise in the bond rate, is equal to the excess of the bond rate over the bill rate. This indifference formula can be expressed as:

$$\mathbb{E}_t[\Delta r_t^L] = \mathbb{E}_t[r_{t+1}^L - r_t^L] = r_t^L - r_t^S, \tag{8.2}$$

where Δr_t^L is the change in the long-term bond rate r_t^L from time t to time $t+1$ and r_t^S is the short-term bill rate.[51] In other words, the expected holding-period returns[52] of bills and bonds should be equal. Furthermore, the expectation of rising interest rates encourages borrowers to borrow long and lenders to lend short, which tends to 'force short-term rates down and long-term rates up' (Stigum, 1990, p. 86).

The idea that expected holding-period returns should be equal, however, is consistent with the expectations theory whereby the current yield curve expresses unbiased expectations of future short rates (Cochrane, 2005b, p. 355), and any combination of short- and long-term assets should yield the same expected return. The theory often includes an additional liquidity risk premium above the expected future rate (Kaldor, 1939, p. 13, n. 1), but, unless this premium is constant, it renders the theory a tautology (Cochrane, 2005b, p. 357). For this reason, the expectations theory fails to find empirical support (Hicks, 1974, pp. 45-46), and studies have shown that implied forward rates are not predictors of future spot rates (Cochrane, 2005b, pp. 426-432).

Crucially, the expectations theory relies on a 'sufficient mass of like-minded persons, all holding the same views with complete conviction' (Kahn, 1954, pp. 230-231). The expectations theory implies homogeneous agents with identical expectations, leading directly to a model populated by a representative agent with Rational Expectations and a neo-Walrasian general equilibrium (Mehrling, 2011, p. 65). As we have seen in Section 6.5, invoking a representative agent with Rational Expectations renders liquidity and money redundant, and Rational Expectations, or identical opinions held with 'complete

[51] Equation (8.2) is original and is intended to be more instructive than accurate, since it ignores the non-linear effect that yield changes have on bond prices, although nothing of importance is lost by omitting it.

[52] The holding-period return of an asset between time t and time $t+1$ is the ratio of the sale proceeds to the asset's purchase price.

conviction', undermine the speculative motive and the demand for liquidity (Kahn, 1954).

Attempts can be made to align the expectations theory with a marginal indifference approach, in which asset prices are set so that the marginal investor, in a body of heterogeneous investors, is indifferent between all assets. These attempts are not entirely successful. Consider an example from Kahn (1954) in which the banking system buys short-term bills from the public by selling long-term bonds. In the real world, according to Kahn, this exchange will cause the bond rate to rise and the bill rate to fall. The change in relative stocks will bring about price changes that maintain the balance between the increased supply to the public of bonds and reduced supply of bills, and a change in the identity of the marginal investor. A greater stock of bonds means that the new marginal investor in bonds must be one that is more bearish than the previous one, with bond rates rising to compensate. A similar argument justifies a fall in bill rates. Under the expectations theory, however, neither should happen, unless the expected future bill rate, which 'predicts' the higher bond rate, has risen precisely as the actual bill rate has fallen, making the expectations theory logically inconsistent with a marginal investor approach.

Thus, the expectations theory of the term structure of interest rates, despite being 'the one that makes the most sense' (Smithin, 2006, p. 279), is inconsistent with the theory of liquidity preference since it directly contradicts its main premise. The determination of interest rates by liquidity preference should be considered in terms of holding-period returns, but not as a prediction of future rates. Liquidity preference is a theory of the long rate of interest and therefore

> …it is misleading to say that the short rate is determined by demand and supply of money while the long rate is determined by the expected future short rate, for one of the main determinants of the demand for money is expectations about the course of the long rate itself. (Robinson, 1951, p. 101, n. 18)

The essential element of organised market liquidity is to allow a limited focus on holding-period returns to only the near future. Organised markets reduce the level of uncertainty about the time required for portfolio adjustment and therefore the strength of the precautionary motive. Expectations could conceivably coincide for the near future, but then deviate as uncertainty causes a divergence of opinion. Prices for both long- and short-term bonds must adjust so that changes in the quantity of money are equilibrated. Thus,

> …the result of increasing the quantity of money is to lower the short rate and to pull the long rate below its expected value to the point where the combined effect of these two movements increases hoards by the amount of the increase in the quantity of money. (Robinson, 1951, p. 101)

Furthermore, Robinson (1951, p. 110) expects that, if the central bank continually lowers its policy rate via open-market operations, then the expected rate of interest will fall, and consequently so will the actual one. The current rate of interest is based on past experience, and the accepted or conventional rate can fall over time, as the bears gradually give up on the idea that rates will rise. This argument conflicts with the idea that interest rates are set purely by uncertainty of their future values, although it conforms well to the apparent results of current central bank policy settings and the low rates associated with QE. Similarly, Bibow (1998, p. 242) hints at the phenomenon that the central bank could buy long-term bonds with new money indefinitely, with a continuously lowering rate as more investors become bears and prefer to hold the new money. This phenomenon relies on a given state of expectations, especially concerning the 'safe' rate of interest, which cannot be guaranteed to remain constant throughout. Beyond a certain point, the rate of interest may not be able to satisfy the liquidity preference and a limitless desire to hold liquidity assets could take hold, with the monetary authority losing control and falling into a liquidity trap. Depending on expectations, liquidity traps can happen at any level of interest rates (Bibow, 1998, p. 243).

The rejection of consistent expectations means that the relationship between liquidity preference and equilibrium needs clarification, especially in light of Fischer Black's assertion that, in equilibrium, asset expected returns should be unaffected by liquidity (see Section 7.5). This clarification can be provided by considering the concept of equilibrium itself. For Hayek (1937), the concept of equilibrium makes most sense in describing the planned actions of an individual, and this equilibrium will last for as long as the individual's expectations prove to be correct. It follows, therefore, that time is an essential element to the concept of equilibrium and that a timeless equilibrium is 'meaningless' (Hayek, 1937, p. 37). Extending equilibrium to cover the interaction of many individuals is problematic, for if each individual creates a plan based on conflicting expectations then their plans cannot be simultaneously executed, and no equilibrium is possible. For a 'social' equilibrium the 'corresponding actions' of other individuals must be compatible (Hayek,

1937, p. 38).

The difficulty in establishing an equilibrium for society as a whole is that each individual will have a different set of subjective facts, so an equilibrium for society can then take one of two forms. In the first form, each individual's plan must be 'mutually compatible' so that there is 'a conceivable set of external events which will allow all people to carry out their plans and not cause any disappointments' (Hayek, 1937, pp. 39-40). This mutually consistent expectations equilibrium 'will continue, once it exists, so long as the external data correspond to the common expectations of all the members of the society' (Hayek, 1937, p. 41). Crucially, the equilibrium itself does not depend on objective data, only on the condition that all individuals have the same expectations, rightly or wrongly, concerning the plans of other individuals and the external facts. In equilibrium, the plans of every individual could conceivably be possible, even if external events subsequently make some or all of them impossible.

In the second form, all individuals have the correct knowledge of the objective data. This form of equilibrium, which corresponds to Rational Expectations, can only be identified *ex post* (Hayek, 1937, p. 40). If we assume that all agents in a model know everything, then a societal equilibrium exists, by definition. Naturally, the subjective expectations of individuals will be based on their observations of the objective data, and the first form of equilibrium can only last as long as the subjective expectations match the objective facts. For Hayek, the only reason to be interested in the concept of equilibrium at all is the 'supposed existence of a tendency towards equilibrium' or that expectations 'become more and more correct' (Hayek, 1937, p. 44), which, in a world without fundamental uncertainty, can be interpreted as Rational Expectations.

The problem of establishing consistent expectations can be sidestepped by the assumption of a representative agent, of course, but this is explicitly rejected by Keynes, for, in the absence of perfect foresight, the stability of the system relies on a '*variety* of opinion' about the future rate of interest (Keynes, 1936, p. 172, original emphasis). The expectations theory of interest rates, with its consistency of expectations, must be rejected when there is more than one individual and a variety of opinions:

> Even if prices are the average of individual expectations, average expectations fail even the basic property of the law of iterated expectations. That is to say, the average expectation today of the average expectation tomorrow of some variable

> is not necessarily the average expectation today of that variable. The normal rules for intertemporal consistency do not apply to average expectations as they do to an individual's expectations. This is why Keynes's beauty contest example – about how the savvy trader is able to go beyond the "average opinion of the average opinion" – is so potent. (Shin, 2017, p. 3)

Sharp asset-price movements are caused by adjustments by individuals that are 'more similar than they are dissimilar in their reaction to news' (Keynes, 1936, p. 199). If all agents were the same, asset prices would be unstable and relative holdings would be indeterminate. Thus, large price movements associated with low volume, rather than being indicative of illiquidity (see Section 7.5), signify a convergence of expectations (Rogers, 1989, p. 196). At the limit, large price movements can occur without any trading as they adjust to new information:

> If the change in the news affects the judgement and the requirements of everyone in precisely the same way, the rate of interest (as indicated by the prices of bonds and debts) will be adjusted forthwith to the new situation without any market transactions being necessary. (Keynes, 1936, p. 198)

The picture drawn is of a market that jumps discontinuously to a 'new equilibrium rate of interest' (Keynes, 1936, p. 199) as expectations adjust to new information, but without necessarily trading at all. As with the Efficient Markets Hypothesis, the flow of information and the consequent reformation of expectations are sufficient to change the interest rate:

> Changes in the liquidity function itself, due to a change in the news which causes revision of expectations, will often be discontinuous, and will, therefore, give rise to a corresponding discontinuity of change in the rate of interest. (Keynes, 1936, p. 198)

Only a variety of opinion will 'cause some realignment in individual holdings of money' because if 'everyone is similar and similarly placed, a change in circumstances or expectations will not be capable of causing any displacement of money whatever' (Keynes, 1936, p. 198). A variety of opinion is required for the demand for money not to be infinitely elastic (Kahn, 1954, p. 247). To prevent large movements into cash after small changes in opinion, each agent must have a different expectation as to the future volatility of the rate of interest. For, if no one felt any uncertainty about the future rate of interest, the speculative

motive would not operate (Keynes, 1936, p. 209), the rate of interest would be determined by real factors (Kahn, 1954, p. 238), and the quantity theorem would hold (Keynes, 1936, p. 209). The greater the similarity there is among investors the greater the move in the interest rate will need to be to restore equilibrium.

The idea that the stability of the system relies on a variety of opinion both conflicts with the consistency of expectations necessary for a REE and introduces difficulties of its own. For example, Boulding (1944) uses the variety of opinion to develop the concept of a 'null price', below which an agent would be willing to add to their holdings of bonds and above which they would sell or issue new bonds. Assuming the continuity of demand and supply, the 'null price' is then on the cusp of selling and buying where the agent would make no changes to their bond holdings. For Boulding, it is divergences in null prices across agents that determine the quantity of an asset exchanged, implying that the divergence is the result of disequilibrium. Taken across all agents, excess demand functions can be constructed similar to those underlying the Arrow-Debreu derivation of general equilibrium and market clearing prices (Arrow & Hahn, 1971). In equilibrium, however, the null prices converge (Wray, 1991a, p. 122; Wray, 1992a, p. 73), thereby contradicting the requirement for a variety of opinion to ensure stability.

Boulding then goes a step further and interprets liquidity preference as a desired ratio of monetary to total wealth, such that increases in the stock of money prompt increases in non-monetary asset prices (or production) to restore the desired ratio to equilibrium. In this way, it is a theory of asset (and commodity) pricing. The quantity traded depends on the divergence between null prices, and the stock of money determines asset prices (Wray, 1991a, p. 120). As in the Walrasian system, interest rates are not determined directly, but are determined by the present price of assets and their future payments (Boulding, 1944, p. 62).

Further complexity arises when integrating the banking system. For a bank, an increase in the null price, or equivalently a decrease in liquidity preference, implies an increase in its propensity to make new loans and issue new bank money, which in turn raises the null prices of the non-bank public (Wray, 1991a, p. 120). An increase in the public's liquidity preference, which, when confronted with a fixed stock of money, would raise interest rates, could conceivably be offset by the banking system (Wray, 1991b, p. 6). In the case of a bearish market, market makers that purchase assets during a sell-off may need to draw on

lines of credit, or funding liquidity, and thereby expand the money supply (Brown, 2003, p. 334). In aggregate, the banking system has the ability to satisfy a degree of liquidity preference, whether stemming from bearishness or not, by purchasing unwanted assets for new bank deposits (Kaldor, 1982, p. 14).

A move by the public into more liquid assets means that liquidity risk has been shifted to the banking system, which must be followed by an adjustment of expected returns on bonds for this to be profitable for the banks. The terms of this transformation are to be found in the interplay between the liquidity preferences of the public and the banking system (Keynes, 1937c, p. 666). First, the banking system must purchase bonds at a decreasing expected return to counter the public's increasing liquidity preference. Second, the creation of monetary liabilities depends on the central bank lender of last resort protection and the banking system's capital structure. This counterbalancing relies on the unlikely event that the banking system's preferences move in the opposite direction from that of the public (Wray, 1991b, p. 6). Instead, increases and decreases are likely to be amplified by the banking system (Wray, 1991b, p. 6).

If liquidity preference falls, the rates on non-money assets fall, the cost of new loans falls and creates the conditions for even more money to be created. Liquidity preference—the desire to hold near-money assets—affects the expected return on all assets. In a bullish market, where every agent prefers securities with a rising price to price-protected assets, 'there is no limit to the rise in price of securities and no effective check arises from a shortage of money' (Keynes, 1930a, p. 256). In other words, a shortage of liquidity cannot stop a bull market. For, in the absence of uncertainty, only the transactions demand for money would operate; there would be no need for precautionary or speculative demand (Runde, 1994, p. 133).

8.6 Stocks, Flows and Asset Fetishism

Underlying each interpretation of liquidity preference discussed so far is an assumption of a fixed stock of assets. 'Prices are then such that the market is content to hold just that quantity of each type of asset which is available at the moment' (Robinson, 1951, p. 96). Liquidity preference shows that the potential supply of liquidity is a fundamental aspect in the pricing of all assets:

> When, as happens in a crisis, liquidity-preferences are sharply raised, this shows

> itself not so much in increased hoards – for there is little, if any, more cash [liquidity] which is hoardable than there was before – as in a sharp rise in the rate of interest, i.e. securities fall in price until those, who would now like to get liquid if they could do so at the previous price, are persuaded to give up the idea as being no longer practicable on reasonable terms. (Keynes, 1937a, p. 211)

On the surface, this passage appears to corroborate the portfolio rebalancing or stock interpretation, with security prices adjusting to equilibrate their holdings. On closer inspection, we can see that all asset prices adjust to the increased demand for liquidity. Liquidity is a systemic factor in asset pricing, hence the appropriateness of the two-asset presentation of liquidity preference. Keynes presents liquidity preference using a static, partial-equilibrium approach designed to show that a less-than-full-employment equilibrium is possible (Brown, 2003). Hence, in this static moment the money 'supply' is one of the elements that are assumed given, since

> …there are two quite distinct types of influence which play upon the equilibrium pattern of rates. One is the state of expectations and the other is the supply of money. To discuss them separately we require that one be assumed constant when the other varies. (Robinson, 1951, p. 98).

Thus, the monetary authority can affect interest rates by two channels: first the supply of money with constant expectations, and second by changing expectations themselves (Bibow, 1998, p. 241). Chick (1983, p. 14), in interpreting Keynes's analytical method, concludes that monetary theory cannot be conducted within either the partial or general equilibrium methods. The former cannot account for the 'macroeconomic repercussions' and the latter has no essential role for money.

The static, partial-equilibrium approach makes no allowances for the dynamics of the money supply and hence the role of banks in supplementing it. Only the finance motive has any role in explaining the dynamics of the system (Bibow, 1995). As such, the finance motive is a disequilibrium concept, since in equilibrium, 'when planned activity equals actual activity' (Bibow, 1995, p. 650), demand for finance is met by a 'revolving fund'. In the transitional period before equilibrium is established, an increase in planned activity requires some exchange of liquidity, therefore 'someone else has to agree to become, for the time being at least, more unliquid than before' (Keynes, 1937c, p. 665). If it confronts a limited supply, this extra demand for liquidity will raise the rate of interest, but not if the

banking system provides the necessary adjustment (Rochon, 1997, p. 288). The finance motive does not support a loanable funds argument due to its disequilibrium nature; the finance motive folds into the transactions motive, once equilibrium is established. Before then the finance motive drives a change in the transactions motive for the planned investor.

> The terms on which finance for any expenditure is available are determined by liquidity preference which thereby affects, most importantly, the volume of the flow of current investment expenditure and saving, the latter being usually obscured by such non-explanatory conceptions as *ex ante* or *ex post* saving or "loanable funds". (Bibow, 1995, p. 663, original emphasis)

An important distinction between Keynes's theory of liquidity preference and the theory of loanable funds is their respective classifications as stock and flow theories (Conard, 1959, p. 218). Loanable funds theory establishes a flow equilibrium of real investment and saving, whereas liquidity preference is a stock equilibrium of money or liquidity (Wray, 1992a). This distinction seemingly allows the pedagogical IS/LM model (Hansen, 1953; Hicks, 1937) to incorporate both liquidity preference and loanable funds, and apparently to reconcile the, otherwise conflicting, issue of equilibrating the stock of money with the flow of the goods market. Because of liquidity preference, the equilibration of investment and saving are achieved by changes income driven by effective demand, not interest rates (Kregel, 1988).

This supposed reconciliation of flow and stock equilibria is a source of confusion as illustrated in interpretations of the finance motive. For instance, Davidson concludes that, in the context of the IS/LM model, the finance motive means that 'demand for money function is *not* independent of changes in the real sector' (Davidson, 1965, p. 52). A conclusion that Bibow (1995, p. 652) translates to be that 'the finance motive would imply that the IS and LM curves are not independent of one another but would generally shift simultaneously.' Bibow finds this conclusion logically inconsistent with the comparative-static equilibrium analysis of IS/LM since it must imply 'a *lasting additional* demand for money', which contradicts the dynamic, disequilibrium nature of the finance motive (Bibow, 1995, p. 652, original emphasis). This error creeps in because 'Davidson defines the demand for money as a function of (planned) expenditure and not, as according to him is usually the case, as a function of income (output)' (Bibow, 1995, p. 652). In equilibrium, it makes no difference whether it is planned or actual expenditure or output; they are all

equal.

The flow of new loans is small relative to the stock of existing loans, so that expectations concerning the value of existing loans are the dominant factor in asset pricing. ‘The price of an asset is determined solely by its expected future price, independently of its current flows of supply and demand, if these flows are dwarfed by speculative stocks that are *very* large’ (Ertürk, 2006, p. 457, original emphasis). Hicks (1989, p. 11) highlights that, in speculative markets where intermediary traders or dealers can hold stocks across trading periods, flow demand and supply cannot determine asset prices in a Marshallian or Walrasian fashion. Futures markets show that prices are not determined by ‘flow propensities’ but expectations, not necessary irrational (Hicks, 1989, p. 17). Instead, a distinction must be made between the exchange of existing assets and the production of new assets in the theory of value (Townshend, 1937, p. 160). Liquidity preference supports this distinction with a ‘two-price’ theory in which ‘asset prices are determined independently of investment and saving flows’, and only are indirectly affected by consumer prices (Ertürk, 2006, p. 463).

Although this idea of the flows being small relative to the existing stocks is decisive, it does not go far enough. Innovation in modern finance has undermined the importance of assets; it has deconstructed assets into risks. Each asset’s risks can be decomposed into its primitive risks via financial derivatives and intermediation, with intermediation that depends on the weights and individual expectations of the public and the banking system. Asset pricing should be seen to be stemming from risk, not stocks, since stocks, if financial assets are included, adjust also to ensure that all risks that exist anywhere are owned somewhere. Nevertheless, liquidity is not represented in this risk framework. The fact that assets can be decomposed, and that derivatives can extend the supply of many of them, moves the focus from assets to risks where the supply cannot sensibly be defined. Just like the QTM which falls apart when the question ‘what is money?’ cannot be answered, the question ‘what is an asset?’ similarly underlines the stock-flow paradox.

The key difference between the principles underlying portfolio rebalancing and stochastic discount factors is whether it is the supply or the riskiness of assets that should determine prices. Black’s (1970a; see also Section 7.5) ontological challenge can be used against the argument for supply and to reinterpret and clarify Keynes’s theory of liquidity preference. Developments in modern finance enable the layering of financial instruments (such as

currency, interest rate, and credit default swaps) over physical assets to disassemble the various risks embedded in the underlying asset. To maintain a focus on assets *per se*, and not their risks and cash flows, is to suffer from a kind of 'asset-fetishism' (Lozano, 2015) and to remain oblivious to the distinction between the analysis of 'classical exchange' concerning 'exchange and immediate settlement of physical assets' (Lozano, 2015, p. 32) and the ontologically broader concepts of generic and synthetic finance. Generic financial instruments are 'non-physical economic objects, whose quality and character of physicality will generally *not* be informative of its value, e.g., loans, bonds, bills and stock', with cash flows that are decoupled from any reference to physical objects so that the 'time-horizon of the tenure of exchange is loosened beyond immediate settlement' (Lozano, 2015, p. 33, original emphasis). Synthetic finance broadens this ontological set further by considering risk-transfer instruments such as interest rate and credit default swaps—instruments that have no 'endogenous limitation' to their creation (Lozano, 2015, p. 38). This ontological insight is crucial to understanding that risk and liquidity are not intrinsic properties of an asset and gives us further reason to reject the commodity view of money.

By arbitrage, asset prices that are not consistent with risk-based models can be undermined by the creation of derivatives, thereby imposing a powerful influence that cannot be ignored. The stock-based argument cannot provide a particularly strong theoretical basis for asset pricing. For example, to claim that an asset's price is 'high' because it is in short supply is, first, to deny the respective roles of systematic and idiosyncratic risk in asset pricing, and, second, to ignore the potential that asset 'shortages' can be mitigated by deconstructing their respective risks via financial derivatives, even if this is not always possible in practice. On the other hand, the risk-based approach relies on the unrealistic assumption that investors can buy and sell assets in any amount they like, prompting questions about the existence of liquidity constraints.

The credit risk associated with intermediaries in synthetic finance is the risk introduced by the process of transformation; for example, maturity transformation or liquidity risk services (Goodhart, 2008). The question is, if financial instruments are created to transfer existing risks, are they economically significant in themselves or can they be safely ignored to focus on primary risks only (Block & Barnett, 2012, pp. 14-15)? Does intermediation create new risks, even as it tries to transfer risks? It can be argued that the creation of financial instruments themselves creates new risks, by becoming new weak links in an

increasing chain of contractual obligations, and that the gross amount of assets and liabilities is important in itself. The chaining of claims across intermediaries must surely increase credit risk, such that the 'scope for introducing risk into the system rather than sharing it around is obvious' (King, 2016, p. 143). That intermediation matters is evident by observing that open-market operations and liquidity provision by banks are economically significant and the dysfunction of the shadow banking system resulted in real economic effects.

For Jefferis (2017, p. 83), synthetic risk transfers allow the universe of assets with 'known' prices to be expanded, with the newly added assets becoming more 'liquid' by means of 'financial calculations [that] do not have veracity because of predictive power but because the practice of financial modelling commensurates different forms of capital in order to make them into a form of liquidity.' That supposedly marketable prices could be generated by pricing models based on risk transfer supported the illusion of market liquidity needed to support repo and the shadow banking system.

The focus on assets and their stocks is misguided. The idea that all assets must be held by someone needs to be adjusted. Financial derivatives make it possible to separate and transfer risks, so the focus must be on component risks: all risks must be held by someone. Just so, Pozsar (2014, p. 66) laments that the existing U.S. Flow of Funds framework omits derivatives thus obscuring the visibility of the flow of risks compared to the flow of funds and hence the understanding of asset prices. Similarly, Brunnermeier et al. (2011) suggest a 'Liquidity Mismatch Index' to improve on basic measures of leverage when applied to derivatives and off-balance-sheet exposures.

The consumption-based asset-pricing model adopts a risk-based approach to asset pricing (see Section 6.6). What is missing from this model, however, is the assessment of liquidity in the process of establishing prices and its relation to the established Wicksellian, intertemporal features in the complete-markets' paradigm. Without recognising the ontological separation of assets from risk, and therefore liquidity, the innovations, developments and consequences of synthetic finance, more commonly called 'financialisation', will remain opaque and misunderstood.

8.7 The Pure Rate of Interest

Liquidity preference is a theory of the interest rate, or 'the pure rate of interest' (Keynes,

1936, p. 208) and asset pricing more generally, but should not be conflated with that of asset prices itself. Instead, it corresponds to the discount factor expressed the generic asset-pricing formula, equation (6.9). The pure rate of interest established by liquidity preference should be contained in the stochastic discount factor, which must therefore be modified to include liquidity preference.

This discount factor versus payoff split is seen in Keynes's analysis of long-term expectations on the uncertainty of prospective yields, isolated from changes in the rate of interest used to discount these yields (Keynes, 1936, pp. 148-149). The distinction between yields and the rate of interest is significant. Uncertainty arising from asset yields creates the incentive to hold liquid assets, which in turn determines the rate of interest. Liquidity preference, or the propensity to hoard, therefore applies directly to asset pricing:

> This, then, is the first repercussion of the rate of interest as fixed by the quantity of money and the propensity to hoard, namely, on the prices of capital-assets. This does not mean, of course, that the rate of interest is the only fluctuating influence on these prices. Opinions as to their prospective yield are themselves subject to sharp fluctuations, precisely for the reason already given, namely, the flimsiness of the basis of knowledge on which they depend. It is these opinions taken in conjunction with the rate of interest which fix their price. (Keynes, 1937a, p. 217)

Note the distinction between the logically separate (Cottrell, 1994b, p. 417; Keynes, 1936, pp. 173-174) processes of fixing the interest rate and forecasting the prospective yield of a specific asset is directly analogous to the asset-pricing formula, $p_t = \mathbb{E}_t[m_{t+1}x_{t+1}]$. The stochastic discount factor m_{t+1}, common across all assets, is the 'pure rate of interest' and x_{t+1}, the asset payoff, is the 'prospective yield', where the units of measurement are not real or percentage returns, but future (and uncertain) monetary flows arising from ownership of the asset. Finally, $\mathbb{E}_t[.] = \mathbb{E}[.|\mathcal{F}_t]$ is the subjective expectation given 'the flimsiness of the basis of knowledge on which they depend' and variations in x_{t+1} will be due to the uncertainty of prospective yields.

This decomposition of the factors of asset pricing corresponds to the 'facts' in the empirical finance literature. Specifically, Cochrane (2011b, p. 1051) reports that that the 'variance in dividend yields or price-dividend ratios corresponds entirely to discount-rate variation' and not forecast changes in dividends themselves. Two sources of variation (or asset-price shocks) are observed: one due to cash flows or dividends and the other from changes in

discount factors, and these shocks are uncorrelated (Cochrane, 2017a, p. 971). Placing these 'facts' into Keynes's framework, we can recognise that the propensity to hoard is the discount factor, and prospective yield aligns with cash flow or dividend yield variations. Keynes's theory allows us 'to think of and model a world with separate cash-flow and discount-rate shocks' (Cochrane, 2017a, p. 971).

This independence of cash flow and discount factor shocks reinforces the rejection of productivity-based theories of interest rates and the Wicksellian notion of a real interest rate determined entirely as a non-monetary phenomenon. It also provides clues as to why conflicting theories of interest rates and prices can exist. On the one hand, we find theories of the rate of interest as a cost to be covered by profits, and hence prices. A lower rate of interest thereby lowers prices (Pivetti, 2009). For entirely different reasons, a similar idea is displayed by the neo-Fisherian idea that higher interest rates increase inflation expectation. By the QTM, however, lower interest rates increase the money supply and hence prices.

Importantly, the speculative motive does not relate asset prices with the quantity of money; it relates the rate of interest with the quantity of money (Keynes, 1936, pp. 173-174). Liquidity preference is not a risk-based theory of bearishness; it is an uncertainty-based desire for liquid assets:

> Whilst liquidity-preference due to the speculative-motive corresponds to what in my *Treatise on Money* I called "the state of bearishness", it is by no means the same thing. For "bearishness" is there defined as the functional relationship, not between the rate of interest (or price of debts) and the quantity of money, but between the price of assets and debts, taken together, and the quantity of money. (Keynes, 1936, pp. 173-174)

Without the recognition that liquidity preference relates to only one of the two factors in asset pricing, there is 'a confusion between results due to a change in the rate of interest [the discount factor] and those due to a change in the schedule of the marginal efficiency of capital [the asset's payoff]' (Keynes, 1936, p. 174).

Ertürk (2006, p. 464) provides an example of this confusion, whereby 'liquidity preference [centred] on the expected variations in the price of loan capital, and delineated the variations in share prices as a separate issue to be dealt with under the marginal efficiency of capital'.

This delineation must be rejected because asset valuation, whether bonds or shares, requires both the marginal efficiency of capital and the discount factor. The marginal efficiency of capital cannot be assigned to shares and the discount factor to bonds, each form of asset requires both to determine their price. Similarly, Minsky (1975, p. 67) believes that Keynes errs by expressing the theory of liquidity preference in terms of interest rates. Minsky evaluates asset prices in terms of discounted cash flows, but his theory explicitly considers the systematic liquidity risk of failing to meet the future cash-flow commitments from current borrowing—the 'survival constraint'[53]. Indeed, Minsky's asset-pricing formula includes not just the asset's payoff, but also the stock of money (Minsky, 1975, p. 102) and the financing arrangements of the investor, with more external funding versus internal funding lowering the demand price of the asset (Minsky, 1975, p. 106), contradicting the Modigliani-Miller theorem. Minsky placed 'emphasis on (non-diversifiable systematic) refinance problems as the source of borrower's and lender's risk, since refinance problems have no place in an equilibrium theory of asset pricing' (Mehrling, 1999, p. 150).

It is important to avoid the misunderstanding that the liquidity premium attached to holding money is conceptually the same thing as the rate of interest. It is incorrect, in the context of the theory of liquidity preference, to claim that 'the money rate of interest refers to the fee from lending *money*' (Hansen, 1953, p. 160, original emphasis). Money and money lending are distinct assets and, in equilibrium, their expected returns, including liquidity premia, should be equal, but they should not be considered the same phenomenon. By example, fees charged by money market funds have been waived since 2008 to avoid negative returns, these fees were reinstated in 2016 so that returns are now negative (Ram, 2016). The visible carrying costs of money market fund investments are outweighed by the abstract liquidity premium associated with their liquidity. Similarly, the negative rates charged by the ECB on bank reserves are a carrying cost, and not a negative liquidity premium.

To clarify, the marginal efficiency of bonds has been lost in most interpretations of liquidity preference, with the assumption that the yield observed on bonds is conceptually the same as the liquidity premium. This is incorrect: the bond price is determined by the liquidity preference-based discount factor applied to the cash flows of the bond. The fact that the

[53] Mehrling (1999, p. 139) highlights this phrase from Minsky's unpublished PhD dissertation.

bond yield is the most directly observable manifestation of the discount rate should not obscure the theoretical point that the discount factor is independent of any particular type of asset. Keynes's 'money market' should be differentiated from the credit market, in which money is lent (Brown, 2003, p. 331, n. 16). The concept of the 'money market' used by Keynes is a fictitious institution and demonstrates that the liquidity premium earned by money is an abstraction that can only be 'earned' in subjective utility terms.

8.8 Summary

This chapter has shown that Keynes's theory of liquidity preference makes a valuable contribution to the analysis of liquidity, although some reinterpretation of the theory is warranted. The meaning of the term 'liquidity' that underpins the theory of liquidity preference can now be located in the taxonomy or spectrum of liquidity identified by this study. It must be at the aggregate, price-protection end of the spectrum, most commonly associated with money and near money. Inasmuch as money (read liquidity) can include the command over general purchasing power that can only be recovered over a period of months, then liquidity preference can be satisfied by quasi-money categories of self-liquidating assets and repurchase agreements. Liquidity preference is not built on market liquidity, and measures of liquidity based on trading volumes, returns reversals, or zero trading days (see Section 7.5) fail to capture the liquidity risk mitigated by assets with a price-protection feature specific to liquidity preference.

The theory of liquidity preference is that agents are willing to forgo interest income to hold price-protection assets due to the capital uncertainty associated with relying on market liquidity. This uncertainty regarding sales price is not just the transactions cost of illiquidity arising from markets without perfect market liquidity; it is due the uncertainty of the discount factor itself. Market liquidity is a fictitious form of liquidity that allows measures of value to masquerade as measures of cash flow.

The interpretation taken in this study is that illiquidity is a systemic property of all assets that do not display the price-protection property, since expected returns are predominantly based on systematic factors. Keynes's two-asset exposition is justified: any further expected return or premium over that determined by liquidity preference is based on the asset's systematic risk, not liquidity. This is not to say that all non-money assets are perfect substitutes, only that the component of the interest rate determined by liquidity preference

is shared by all assets. Liquidity preference then becomes a theory of asset pricing.

The characteristic of full value, outside of general equilibrium theory, begs the question. If money is non-neutral, no long-period neo-Walrasian equilibrium exists, and the concept of a fundamental value is undefined. We must eschew the idea that ‘full value’ represents ‘without loss’, full or fundamental value relies on a general equilibrium, and perfect market liquidity, which has no requirement for money, and is not relevant to the consideration of liquidity. It is the mistaken path to considering liquidity as a friction.

Having discussed and interpreted liquidity preference from the ontological perspective of this study, in the next chapter we consider the post-Keynesian contributions to Monetary Analysis: the theory of the monetary circuit and the theory of endogenous money.

9 Post-Keynesian and Heterodox Monetary Theory

9.1 Introduction

Post-Keynesian monetary theories explicitly consider the role of the banking system in the operation of an economy where money and credit have a fundamental place. The credit nature of money is recognised and essential to each theory. As with the theory of liquidity preference, the point of departure for post-Keynesian analysis is a rejection of the axiom of money neutrality. Instead, money has real effects. Post-Keynesian theories emphasise the process of money creation, and eschew the idea that money is an external phenomenon with the sole purpose of establishing price determinacy.

Post-Keynesian theory as a whole also shuns the deductive and instrumentalist approaches prevalent in Real Analysis. The quest for microfoundations is deemed inappropriate for macroeconomic analysis (King, 2012), as is general equilibrium theory (Jespersen, 2009). These methodological differences allow liquidity and money to appear as social phenomena, and for uncertainty to display its natural characteristic of increasing over time, rather than dissipating, as in general equilibrium theory.

Most important, the theories presented in this chapter adopt a different stance on the liquidity abundance versus liquidity scarcity issue. In general, the liquidity abundance view is embraced, and hence the limits to pure credit expansion are less clear. Although none of these theories addresses liquidity directly, a brief summary of each is provided in this chapter because each has important elements necessary for the synthesis developed in the next chapter. Each theory has important insights to be identified and examined.

9.2 Circuit Theory

The credit view of money can be refined by the theory of the monetary circuit with the assertion that money ‘is the by-product of a balance sheet operation of a third agent [a bank]’ and ‘*always* emerges as a debt (or liability) issued by this third agent on itself’ (Parguez & Seccareccia, 2000, p 101, original emphasis). Circuit theory revives the terms of the classical Banking School and combines them with the metaphor of a circuit to describe the process of efflux and reflux of bank credit (see Section 4.3). In logical time, the circuit begins with banks’ enabling firms to fund the process of production by an exchange of credit between the firms and the banking system (Graziani, 1990). Households

earn wages paid by the firms in the form of bank credits, which are then used to purchase the goods and services produced or, alternately, new financial securities issued by the firms. Finally, firms can then repay their bank loans and the circuit reaches closure.[54] Circuit theory highlights the importance of the distinction between firms, banks and households in the creation and circulation of money and the determination of interest rates (Seccareccia, 2012).

If households retain savings in the form of bank-money balances, then firms will remain indebted to the banks, thereby preventing the closure of the circuit (Fontana, 2000, p. 35). In response, the business sector can issue financial assets that pay sufficient interest for the household sector to exchange them for their deposits, which can subsequently be used to repay business loans from the banks. The interest rate on the financial assets compensates the household sector for this reduction in liquidity. Since 'money balances measure the outstanding debt of firms towards banks', the household sector's desire to hold money results in higher interest rates and restricts the ability of 'firms to renegotiate a new flow of money' (Fontana, 2000, p. 37). Hoarding 'obstructs the process of wealth creation upon which the value of money depends', since money 'has a value only as long as it is spent by nonbank agents for the purpose of creating future wealth' (Parguez & Seccareccia, 2000, p. 105).

Circuit theory represents a significant reinterpretation of the classical velocity of money by describing the process of money creation as part of the production process in a credit economy in which money is created as a by-product of the 'income-generating finance process' (Lavoie, 1984, p. 788). Instead of the classical metaphor of the same exogenous stock of money circulating throughout the economy, the circuit consists of the continual creation and destruction of credit (Rochon, 1997, p 286). Rather than a fixed scarcity of liquidity, the circuit implies the continual creation of liquidity by the banking system to meet the needs of trade. The amount of lending by the banking system at the beginning of the circuit is determined by the factor costs of the producers. Credit requirements arising from the production process result in the creation of bank deposits. An implication of the circuit theory is that the supply of, but not necessarily the demand for, money depends on income.

[54] See Passarella (2014, p. 130) for a visual depiction of the process.

Money is created during the production process to pay income or factor costs, in other words, the value of output:

> The consumers' purchasing power is therefore largely supplied out of the credits which the traders borrow from the banks. Credit originates in production and is extinguished in consumption. The supply of purchasing power is thus regulated by the transactions which require to be financed. (Hawtrey, 1919, p. 10)

It follows that any limitation on borrowing for production, is a limitation on income and expenditure. Circuit theory speaks directly to trade liquidity (see Section 5.4), where both consumers and producers are justified in expecting liquidity due to the very nature of the credit and banking system that supports their activities.

The flow of funds through the circuit is a dynamic, disequilibrium process. Only after income earners have satisfied their desired purchase decisions (first reflux) and then their desired mix of securities and money (second reflux) is a recognisable state of equilibrium achieved. By comparison, a neo-Walrasian general equilibrium is conceptually equivalent to the moment of circuit closure, with the added qualification that there is no residual demand for money. As noted in Section 7.4, a centralised, Arrow-Debreu market can be re-interpreted as a decentralised market where transactions are conducted using trade credit, which is subsequently cleared. The Walrasian auctioneer conducts a process whereby 'multilateral trades are possible through a clearing operation that keeps track of net claims' or settlements (Ljungqvist & Sargent, 2004, p. 214). In effect, the Arrow-Debreu market achieves the form of final settlement described in Section 4.8, where all financial debits and credits are extinguished. In equilibrium, neither money nor financial securities exist, and a credit intermediary such as the banking system is unnecessary. The key element missing from neo-Walrasian analysis is an equilibrium in which financial securities remain in existence.

Financial securities are problematic for both circuit theory and neo-Walrasian theory. The development of the shadow banking system poses some challenges to circuit theory with its specificity to the flows of production and consumption. The rise of 'financialisation', which 'refers to a process in which the financial markets have taken on a central role in bank-based economic systems' (Seccareccia, 2012, p. 282) has heralded a dramatic increase in the ratio of financial assets and liabilities to GDP (Michell, 2017). Households, traditionally net lenders in the monetary circuit, have significantly increased borrowing for

house purchases (Sawyer, 2013). Furthermore, securitisation, whereby banks originate loans and then package them into a marketable form to distribute to investors, is, for many, the prime cause of the GFC, with the conclusion that securitisation should be prohibited (Davidson, 2008; Seccareccia, 2012).

Circuit theory is not naturally expressed in terms that provide insights into non-traditional bank lending for the purchase of non-investment assets. For Michell (2017, p. 373), applying circuit theory to the creation and final finance of mortgage assets is a 'perverse and mutated form of the monetary circuit' since it loses its original relevance to the financing of production. Circuit theory needs to be updated for the developments associated with financialisation (Passarella, 2014). For Lysandrou (2014), circuit theory is unsuccessful in describing financialisation because it lacks asset management as a separate sector in its own right. Seccareccia (2012) invites us to believe that, because of financialisation, households have moved from being net lenders to net borrowers, with the corporate sector now taking the position as net lenders. The view that households can be net borrowers, however, does not take into account the observation that the corporate sector is ultimately owned by households. Instead, some households must be debtors to other households via the wealth ownership intermediated by the corporate sector.

Michell (2017), using the circuit theory as an analytical framework, makes the persuasive case that the traditional banking system must still be the source of new money, even for the shadow banking system. Financialisation describes the process in which the financial sector has evolved to offer new liquid products. Due to a desire by households to hold liquidity, the circuit ceases to close, and this closure is prevented by means of new types of financial products. Securitisation, then, is a new form of final finance (Botta, Caverzasi, & Tori, 2015) and one significant provider of these new products is the shadow banking system. Similarly, shadow bank liabilities are not money from the perspective of the circuit theory because they do not function as a medium of exchange; instead, they represent final finance (Gabor & Vestergaard, 2016; Michell, 2017).

The theory of monetary circuit provides a valuable analytical basis with which to understand the process of inside money creation and a means of interpreting the (hidden) dynamics behind the Walrasian auction and general equilibrium.

9.3 Endogenous Money

The post-Keynesian theory of endogenous money also aligns with the credit view of money and has significant implications for causality in the money creation process (Hewitson, 1993). Under a commodity view of money, the money creation process cannot be influenced by the real economy and is theoretically exogenous, but with a credit view, money creation is an essential part of commerce and is theoretically endogenous (Moore, 1988, p. xi). Unlike the QTM (see Section 3.5), which relies on the extension of the exogeneity principle to fiat money, endogenous money theory builds on the basic insight that 'loans create deposits', but not by means of the money multiplier (see Sections 4.6 and 4.9). Instead, endogeneity reflects the notion that the money supply—a concept that is itself difficult to define—is determined by the process of production and commerce, '*because* commercial banks can create credit and money *ex nihilo*' (Rochon, 1997, p. 282, n. 6, original emphasis). In terms of the equation of exchange, equation (3.1), the theory of endogeneity represents 'reverse causality', whereby money is created to finance the value of output.

One of the key differences between endogenous money and the QTM is whether the supply and demand of money are independent (Le Bourva, 1992), and whether the stock of money then resembles a 'hot potato' which gets passed from hand to hand, because the economy 'as a whole cannot get rid of it' (Tobin, 1963, pp. 408-409). In the view of the QTM, there is no reflux outlet for fiat money and any excess money must affect prices or income as the excess moves through individual portfolios until it finds a home (Friedman, 1956).[55] The theory of endogenous money rejects this attempt to treat money like a commodity with a distinct and independent production function, instead it examines 'the extent that changes in the supply of money are caused by changes in the demand for money' (Bibow, 2000, p. 533).

The post-Keynesian monetary literature contains detailed research into the workings of the modern financial system, and the technicalities of the banking system (Fullwiler, 2003) giving it a solid ontological grounding. This literature emphasises that overdraft arrangements commit banks to fulfill pre-arranged lending obligations and, consequently,

[55] When coupled with the axiom of money neutrality, the income effect can only be short term.

the total amount of lending is determined by borrowers' demands for bank money at the prevailing complex of interest rates (Fontana, 2004). When unused overdrafts are considered, the potential amount of lending—and hence the potential money supply itself—is greater than that measured by the current stock of outstanding bank deposits. Furthermore, the potential amount of lending is not constrained by central bank reserves for the simple reason that the central bank must supply the reserves required by the banking system to ensure that the interbank rate matches its target policy rate. Failure to provide the reserves necessary to validate the amount of lending would, at best, drive the interbank rate away from its target, and, at worst, violate its responsibility to prevent financial instability (Wray, 2006, p. 273).

Endogenous money theory can also be seen as an extension of the theory of the monetary circuit (Rochon, 1999). The existence of bank money created at the beginning of the circuit entails an equivalent borrower who must regain bank money to repay the loan, thus setting up a strong tendency for the bank money to return to the borrower, either by sale of goods or new financial securities. The difference between the amounts of bank money borrowed and repaid leaves the stock of bank money determined as a residual and not a causal factor (Davidson & Weintraub, 1973).

> In talking of the "stock of money," we are yielding to convention and habit of mind. The money stock is in fact the resulting factor of the expansion of credit. It can explain neither employment nor prices. The money stock is a residual and as such it cannot be causal. (Lavoie, 1984, p. 775)

Credit is created and cleared by the banks in a process involving an exchange of producer's credit for bank credit. Bank money is the predominant medium of exchange, which means that 'the volume of nominal lending *by* the banking system determines the volume of nominal lending *to* the banking system' (Moore, 1988, p. 62, original emphasis).

Most significant is the rejection of the money multiplier as a deterministic mechanism for controlling the broad money supply (Le Bourva, 1992). Rather than the money multiplier hydraulically extending the base money supply, the endogenous theory stipulates that it is the demand for credit that creates the broad money supply, with the required reserves of base money created afterwards. Central banks ultimately control the amount of their liabilities they create, but in practice, and because of convertibility, the desire for an orderly market means that they allow for the elastic creation of base money. Importantly, because

the majority of the banking system's assets have limited market liquidity or marketability, the central bank has only a limited ability to reduce the monetary base without dire consequences for the stability of the banking system. Conversely, the central bank can create new money by means of open-market operations, so that some aspect of the base money supply can be seen as exogenous (Moore, 1988, pp. 15-16). Overall, the central bank's control over the monetary base is asymmetric (Moore, 1988, p. 23).

For reasons of financial system stability, therefore, central banks have no alternative but to use interest rates as their policy instrument. By necessity, they operate to achieve a target interest rate in the interbank or repo markets. The central banks that operate a channel system aim to keep the stock of reserves close to zero, even though reserves must be available to banks for intra-day settlements to maintain the stability of the payments system. The channel system explicitly uses interest rates to limit the growth in the stock of broad money (see Section 4.6).

The dependency of the bank money supply on credit demand leads to claims that bank money lacks traditional supply and demand curves, since the 'supply and demand for credit money are not independent' (Moore, 1988, p. 68). To claim, however, that the money supply is 'demand-determined' is to conflate three separate phenomena: the demand for credit; the demand for bank liabilities or credit money; and the demand and supply of base money. Much confusion in the endogenous money literature stems from the failure to maintain clear distinction between credit and credit money (Rochon, 1997, p. 281). First, the demand for credit relies on the credit worthiness of the borrower and the cost of borrowing specified by the banks. Without a pre-arranged line of credit, any attempt by a borrower to obtain credit will not necessarily be satisfied by the banking system (Dow, 1996; Kahn, 1954, p. 254; Wray, 1992b). Instead, credit restrictions are applied based on factors other than the prevailing rate of interest, and such quantitative limits mean that there is 'normally a fringe of unsatisfied borrowers' (Keynes, 1930b, p. 365).

Second, the demand and automatic supply of credit via an overdraft leads to equivalent creation of bank credit money, which is accepted as a medium of exchange, and becomes 'convenience lending' to the banking system (Moore, 1988). Since money is always accepted in exchange, Moore (2001, p. 18) asserts that the 'general acceptability of money in exchange is independent of the subjective value of the real services-in-kind on money balances, whether this is expressed as an own rate of return, or as the opportunity cost of

holding money balances.' Ignoring the demand for bank deposits, however, renders endogeneity theory a form of 'reverse monetarism', in which depositors are willing to hold whatever amount of money the banking system creates (Bibow, 2001, p. 606). The willing acceptance of money does not necessarily imply a desire to hold money (Arestis & Howells, 1996). Recall the distinction between the flow value of money and its stock value (see Section 7.2). Money can be created, on demand, to meet the needs of trade, but it does not necessarily follow that the residual stock of money is optimal (Goodhart, 1989, 1991; Howells, 1995, 1997).

The question of whether bank money is a 'hot potato' itself reveals a vision of the demand for money as the simple transposition of the supply curve of commodities viewed through the lens of classical exchange and perfect barter (Moore, 2001, p. 17). Confusion in the endogenous money literature about the possibility of an excess supply of deposits created by endogenous money overlooks the important point that banks must pay market interest rates on their non-money term deposits (see Section 4.4). The endogeneity of bank money stems from the banking system's commitment to convertibility into state or central bank money. Convertibility is managed by the interest rates banks pay on non-convertible liabilities with 'the rate of interest as being determined by the interplay of the terms on which the public desires to become more or less liquid and those on which the banking system is ready to become more or less unliquid' (Keynes, 1937c, p. 666).

Crucially, Moore (1988, p. xiii, n. 9, original emphasis) states that 'for analytical purposes the credit money supply must be broadly defined as *all liabilities* of financial institutions issuing transactions deposits.' This definition of credit money includes non-convertible, and hence non-money, deposits bearing a market rate of interest matching the supply and demand of all funding sources (Kaldor & Trevithick, 1981, p. 13). Similarly, Kaldor (1982, p. 46, original emphasis) asserts that an excess supply of credit money '*could never come into existence*', since 'banks provide easy facilities to their customers for switching balances on current accounts into interest-bearing deposit accounts, or vice versa' (Kaldor, 1982, p. 14). There cannot be a surplus of bank money, since if

> …more money comes into existence than the public, at the given or expected level of incomes or expenditures, wishes to hold, the excess will be automatically *extinguished* – either through debt repayment or its conversion into interest bearing assets – in a way in which gold could not be made to disappear from existence

> merely because particular persons find they have too much of it. (Kaldor, 1982, p. 22, original emphasis)

Excess money deposits are converted into non-money deposits by means of the adjustment of deposit interest rates as part of liability management (Lavoie, 1999). The non-money deposit rate must satisfy the public's preferred mix of demand and term deposits. The overall cost of funds is determined by the banking system's behaviour with regard to their preferred mix of monetary and non-monetary deposits—the liquidity preference of the banking system. The demand curve for depositors is therefore upward sloping in interest-money space (Moore, 1988, p. 61). If loan interest rates are set as a mark-up over funding costs, then the issuance of credit money depends on 'the interest rate charged for bank loans and the corresponding interest rate paid on bank deposits' (Moore, 1988, p. 20).

Third, with regard to the demand and supply of base money, the extension of the endogeneity principle to state or central bank money is more problematic. The difficulty is that, for some observers, it is self-evident that the 'community cannot get rid of its currency supply' (Tobin, 1963, p. 415). The evidence for this view stems from a misrepresentation of the gold standard applied to fiat money, whereby it is mistakenly believed that the supply of money under a gold standard is determined by the supply of gold. As seen in Section 3.4, the money supply is also endogenous under a gold exchange standard (Glasner, 1985; Rochon & Rossi, 2013). As established in Chapter 3, all money—including fiat money—is credit and therefore has the ability to be extinguished via reflux or final finance. For example, the current practices of managing a fiat currency are derived from the practices used to manage a gold standard. Under a gold standard, the value of money is determined by the value of gold, but the preservation of this link to gold is performed by Bank Rate, which provides the reflux mechanism required to maintain sufficient reserves of gold. This Bank Rate framework now provides the practical anchor for fiat money.

Although fiat money appears irredeemable, it can be refluxed by various means. Central bank operations require short-term asset purchases and sales, all of which mature and provide opportunities for banks to adjust the amount of reserves they hold. Banks can refuse to sell or repo assets to the central bank during open-market operations thereby preventing the creation of new fiat money. At the end of the repo term, banks can return the central bank reserves and receive the asset. The central bank must then re-conduct all maturing repo agreements to maintain its monetary policy stance. Any reluctance on the part of banks

to enter new repo agreements would be a form of reflux—a preference for assets rather than central bank reserves. The central bank must remove reserves from the banking system when its policy rate has fallen below target. As long as central banks target interest rates as their policy, reflux mechanisms must exist and fiat money must be redeemable.

When fiat money is understood as an evolution of money once convertible to gold, and always managed by way of Bank Rate, it is not difficult to see that the central bank's management framework has not changed. The amount of fiat money in existence is based on that which is demanded at the particular level of Bank Rate, in exchange for the appropriate collateral. The value of fiat money, which is no longer pegged to a commodity, is less obvious. Once the control of the broad money supply by means of the base money supply is rejected, the causal link between price inflation and base money is rejected as well. Nevertheless, when a central bank adopts an inflation-targeting regime, the public can anticipate central bank policy and the expected future path of interest rates is in some sense 'known' or endogenous.

The more contentious claims of money endogeneity concern the degree of freedom that central banks have to set their policy rate. Unlike New Keynesian models, with integrated Taylor rules (see Section 6.3), post-Keynesian models allow the central bank to set the rate of interest anywhere within a broad range (Moore, 1988, p. 266) with a variety of recommended policy rules (Rochon & Setterfield, 2007). This freedom of manoeuvre can be justified empirically by the near-zero, or even negative, rate policies enacted by many central banks since the GFC (Rochon & Rossi, 2013, p. 212) and supported theoretically by a rejection of both the Wicksellian natural rate of interest and the neutrality of money (Rogers, 2006a). Endogenous money theory contends that the interest rate is the central bank's policy instrument, even if this lever is applied by controlling access to base money.

The central bank's policy rate is exogenous in the control sense (Wray, 1992b, 2006) and somewhat exogenous in the theoretical sense because the central bank is free to choose its own reaction function. For example, with constant expectations, the traditional LM curve in interest-money space can be visualised as being horizontal at the central bank's policy rate. Over a longer time-horizon, and if the central bank reaction function is defined in terms of the base money supply, the LM curve could be conceivably be represented as upward sloping, without violating the principles of endogenous money (Fontana, 2004; Wray, 2007). Without considering the implications of the time-frame of the analysis, the

'horizontalist' or 'accommodationist' view can appear to be at odds with the more long-term 'structuralist' view (Dow, 1996).

Endogenous money theory, like the theory of liquidity preference, explains that interest rates are determined independently of the income generating process and therefore 'do not necessarily increase during economic expansions' (Rochon, 1997, p. 277). Central banks do, however, have the ability to adjust their policy interest rate a level that 'replicates the classical results' to balance saving and investment (Moore, 2001, p. 28). The fact that Bank Rate is the policy lever has complications for the theory of liquidity preference. The banking system does not need to be a passive participant in the process of bank money determination. In the face of weakening credit demand banks can try to maintain their profits by purchasing non-loan securities to maintain their balance-sheet size. Bibow (2000; 2001) identifies active balance-sheet management as a form of exogenous money in the sense that it is neither under the control of the central bank or the public, but is determined by bank behaviour. Significantly, if banks were to maintain a constant money stock in the face of falling economic activity, then the 'real balance effect' (Patinkin, 1965), 'Keynes effect' (Cottrell, 1994a) or 'Keynes mechanism' (Bibow, 2000, 2001) would move interest rates in the direction predicted by the loanable funds theory (Bibow, 2001, p. 607, n. 1). This effect is contingent on the behaviour of the banking system and is therefore by no means an automatic effect (Bibow, 2001, p. 610). In responding to a recession-induced rise in liquidity preference by the public, banks would have to raise their term deposit rates to maintain their target liquidity ratio (Wray, 1992b, p. 304). Consequent rises in loan rates would deter new lending and lower the stock of bank money. Any Keynes or real balance effect would be undermined.

Many post-Keynesians (Moore, 1988; Rochon, 1997) argue that the endogeneity of money undermines, or at least necessitates a reinterpretation of, Keynes's theory of liquidity preference, with the short rate exogenously determined by central bank policy and long rates derived from expectations of future central bank policy (Moore, 1988, p. 259; Moore, 1994, p. 122). As seen in Section 8.5, the expectations theory of longer-term interest rates lacks empirical support and is indeed inconsistent with liquidity preference.

> The endogenous money approach rightly emphasizes that monetary policy sets the short-term rate of interest, thereby influencing effective demand and potential growth – the neutrality of money argument is never taken seriously in Post

> Keynesian models. But little is said about how setting the short-term rate relates to financial markets and interest rates in general (other than "mark-ups"). Banks appear strangely passive in this approach and liquidity does not seem to play any significant role, while little is said about policy communication either. (Bibow, 2009, p. 28)

The recognition of liquidity preference as a theory of asset pricing, and not a description of the demand for money, provides avenues for its reconciliation with endogenous money (Brown, 2003; Chick & Dow, 2002; Smithin, 2017). As such, the theory of endogenous money can be viewed as both a complement and an extension of Keynes's theory of liquidity preference. Coupled with the insights of money endogeneity, liquidity preference provides a theoretically sound and practical alternative to the theory of loanable funds (see Section 6.2) as a guide to monetary policy (Bibow, 2005; Tily, 2006). Whereas the Wicksellian notion of the natural rate of interest is deficient:

> …the liquidity preference theory of interest is plainly a piece of pure logic, filling the gap, but in itself leaving some important questions unanswered. In particular, the liquidity preference theory of interest neither explains why the general public's *propensity* to hoard is what it is any time, nor why the banking system provides a certain amount of liquidity at any time, and neither more nor less. (Bibow, 2001, p. 611, original emphasis)

Central banks 'set the overnight rate' but not the 'interest rates which liquidity preference theory is concerned with' (Bibow, 2001, p. 606). The theory of liquidity preference explains the premium between long- and short-term rates, and the short rate itself is determined by the relationship between bills and the 'active transactions balances' of money required to satisfy the medium of exchange function (Wells, 1983, p. 527). The clarity that Keynes fails to provide 'between his basic proposition that the amount of money demanded is inversely dependent upon the rate of interest and the completely different proposition that the equilibrium rate of interest is inversely dependent upon the amount of money' (Patinkin, 1965, p. 372) is resolved by endogenous money theory. The amount of money supplied and demanded is dependent on the rate of interest, which is anchored by the central bank.

That the central bank's policy rate represents the marginal cost of bank funds (Moore, 1988, p. 59) is especially apparent in a channel system. As long as they have the required collateral, commercial banks are able to draw down their lines of credit at the central bank,

at the rate set by the central bank. The relationship between the commercial banks and the central bank is a mirror of that between the public and the commercial banks, just a step higher in the hierarchy of money. 'The liquidity preference of the nonbank public can be satisfied by the stock of bank liabilities, while the liquidity preference of banks can be satisfied by high-powered money' (Wray, 1992b, p. 301). For example, in a system based on the supply of reserves, if the central bank restricts the amount of reserves available to the banking system, banks will need to lengthen their deposit maturities to reduce their need for liquid reserves. By raising the rate of interest on term deposits, they encourage depositors to move out of demand deposits. This higher cost of funding increases loan rates, thus reducing the demand for credit, credit money and hence reserves. The banking system adjusts to a change in monetary policy by adjustments on both the demand and supply sides of credit. Because of the dependence of the supply of credit money on the demand for credit, rising interest rates have a tendency to reduce the money supply.

The theories of liquidity preference and endogenous money are, to a certain extent, products of their time. To reconcile them requires placing them in their respective contexts in the evolution of the financial system.

9.4 Evolution of Finance

Both the theory of liquidity preference and endogenous money need significant re-interpretation to maintain their consistency with the evolution of the financial system. To see this, consider the provision of price-protection assets under an 'auto-economy' compared with an 'overdraft economy' (Hicks, 1974). In the former access to liquidity depends on the 'actual possession of liquid assets' (Hicks, 1974, pp. 50-52), whereas in the latter, no liquid reserves are kept and liquidity is obtained on demand from banks (Hicks, 1974, p. 54). Although a mixture of the two occurs in practice, Mehrling (2016) traces an evolution from a predominantly auto-economy, immediately post WWII, to a predominantly overdraft-economy, and labels the basis of each system as 'Monetary Liquidity' and 'Funding Liquidity', respectively.[56] Mehrling (2016) then postulates that the

[56] The re-use of terms from the taxonomy of liquidity in Chapter 5 adds the potential for confusion in this section. To avoid this confusion, Funding Liquidity and Market Liquidity, introduced here as the basis of monetary systems, will begin with capital letters. The elements of the taxonomy, funding liquidity and market liquidity, are in lower case.

development of the shadow banking system has heralded a new period, one based on 'Market Liquidity', where asset sales or collateralised borrowing are the predominant means of accessing liquidity. Post GFC, the introduction of QE and financial regulatory reforms have led to a reversion to a system based on Monetary Liquidity, and the impotence of monetary policy in the post-war period, has reappeared. In addition, regulatory changes are reducing dealer balance-sheet capacity and, hence, market liquidity.

Each monetary system evolved in response to the contemporary situation. In the system of Monetary Liquidity, the 'private sector as a whole was plentifully supplied with financial assets and, in the early post-war years, was under little pressure to borrow in order to obtain funds' (Radcliffe Committee, 1959, p. 15). The difficulties that the monetary authorities had in limiting credit expansion in the post-war Monetary Liquidity period were ameliorated by the transition to the Funding Liquidity period. After the post-war surplus of financial assets was reduced, a reliance on borrowed liquidity became the norm—a system based on Funding Liquidity and liability management (Moore, 1988, pp. 26-36). The 'Keynesian' interpretation of monetary policy ineffectiveness was replaced by the power of the banking system using Funding Liquidity under an overdraft system (Hicks, 1974, p. 56; Jefferis, 2017), a power then undermined by the Market Liquidity system. Modern central bank channel systems align more closely with a Funding Liquidity system and the overdraft economy (see Section 4.6), in which the endogeneity of money is more clearly displayed.

The monetary literature often assumes that the banking system is based on Monetary Liquidity, either in theory (Patinkin, 1965) or historically (Hicks, 1974, p. 55). Just so, some post-Keynesian authors identify monetary exogeneity in historical periods where commodity money, and not credit money, supposedly prevailed (Chick, 1992; Moore, 1988; Rogers, 1989). Chick (1992) describes hypothetical stages in the evolution of the banking system, beginning with Stage 1 where banks are simply repositories of the public's savings in state money. Stages 2 and 3 broadly correspond with Monetary Liquidity, and Funding Liquidity is realised in Stage 4. The more recent development of Market Liquidity is not represented in Chick's framework. In this 'evolutionary view', then, the endogeneity of money is a recent phenomenon made possible by central bank support and financial innovation. This study, however, has adopted the ontological view that all money is credit (see Chapter 3), and no such concession is necessary. Instead, the credit nature of money

supports the ‘revolutionary view’ that money is endogenous both logically and historically (Rochon & Rossi, 2013, p. 216). The evolutionary view suffers from the ‘loanable funds fallacy that savings deposited with banks may somehow disappear in the banking system’ (Bibow, 2001, pp. 605-606).

The theory of liquidity preference was primarily developed to interpret a system of Monetary Liquidity. The precautionary motive, for which liquid assets are held, corresponds most closely to Monetary Liquidity. The theory can also be used to interpret the subsequent phases. The reliance on funding liquidity and the commonplace access to overdrafts, lines of credit, and credit cards have reduced the need for dedicated balances of cash to satisfy either the transactions or precautionary motives. For any level of economic activity, the amount of money required to satisfy these motives has fallen. Two phenomena, more prevalent in the modern environment, affect the interpretation of liquidity preference by operating in conflicting ways on the motives for holding cash balances. One relates to the increasing availability of liquidity via overdrafts, the other concerns the rate of interest paid on liquid assets.

First, the increasing prevalence of overdrafts and credit cards is a challenge to the motives for liquidity preference, which

> …partly depend on the cheapness and the reliability of methods of obtaining cash, when it is required, by some form of temporary borrowing, in particular by overdraft or its equivalent. For there is no necessity to hold idle cash to bridge over intervals if it can be obtained without difficulty at the moment when it is actually required. (Keynes, 1936, p. 196)

The strength of the transactions and precautionary motives depends on institutional arrangements: if money creation is ‘demand driven’ then these two motives are significantly weakened. To understand the implications requires an analysis of the supply of liquidity by the banking system, since the confidence that agents have in this source of funds determines the strength of the transaction and precautionary motives. This confidence, which will vary with systematic factors, will become difficult to distinguish from changes in the discount factor stemming from the speculative motive in a crisis.

Second, it is now common for deposits at commercial banks and central banks to have non-zero rates of interest. This phenomenon presents a counter-effect, for ‘[i]f deposit interest

is earned or if bank charges are avoided by holding cash, this decreases the cost and strengthens the motive [for holding transactions and precautionary balances]' (Keynes, 1936, p. 196). If liquidity is readily available then there is a disincentive to hold liquidity balances, but this disincentive is offset by the reduced opportunity cost of the non-zero rate of interest paid on liquid hoards.

Going further, however, the shadow banking system, with its reliance on market liquidity, has prompted a subsequent reinterpretation of central bank operations. The innovation of the shadow banking system has increased the strength of the speculative motive. As the reliance on funding liquidity has shifted to market liquidity, so has the nature of the backstop that the central bank must offer. Central banks have morphed from lender of last resort to dealer of last resort: the ultimate provider of, not funding liquidity, but market liquidity (Mehrling, 2011). Similarly, the ultimate backstop to the international system of Monetary Liquidity, the IMF (Hicks, 1974, p. 55, n. 14), has been replaced by a network of central bank currency swap lines (Mehrling, 2015), each providing overdraft or Funding Liquidity to the international system.

As discussed in Section 8.1, Keynes's speculative motive for liquidity relies on an organised market. One such organised market is the one maintained by the active monetary management of the central bank. The central bank sets the rate of interest as a dealer in central bank money and this rate, compared to each individual bank's safe rate, determines the amount that the central bank must provide: 'the banking system and the monetary authority are dealers in money and debts and not in assets or consumables' (Keynes, 1936, p. 205). Without this automatic provision of central bank reserves, as, for example, under a gold standard, the precautionary motive and the holding of actual reserves would be the more prevalent form of liquidity management.

If the central bank restricts this organised market to short-term bonds only, then the interest rates on long-term bonds are market determined. The central bank could extend their market making to bonds of all maturities, as was attempted in the immediate post-war period (Tily, 2006). If the intention were to push the rate of interest on longer bonds below the 'safe' level, then the result could be that the public refuses to own the longer bonds at all. The long-term rate 'may be more recalcitrant when once it has fallen to a level which, on the basis of past experience and present expectations of *future* monetary policy, is considered "unsafe" by representative opinion' (Keynes, 1936, p. 203, original emphasis). For

example, under a gold standard the 'safe' level of interest is given by the rate prevailing in the international market, and a domestic rate that deviates too much from this level will be 'viewed with a justifiable lack of confidence' (Keynes, 1936, p. 203). In the end, 'the monetary authority would have lost effective control over the rate of interest' with the result that 'the public authority itself could borrow through the banking system on an unlimited scale at a nominal rate of interest' (Keynes, 1936, p. 207).

This picture conjures up images of the present practice of QE, but there are important differences. First, in conducting QE, the authorities have not declared a target rate of interest; they have generally declared a target amount (in monetary terms) of purchases. This has left a degree of uncertainty in the market as to what the interest rate should be, and has, to a certain extent, undermined the policy itself. Second, the picture Keynes paints is of a market whereby ownership of financial assets has been transferred in its entirety onto the balance sheet of the monetary authorities because the public prefers cash or bank liabilities. Since all trading in interest-bearing assets has effectively ceased, the interest rate is undefined. What is not clear in this analysis is the effect on equities and other types of financial asset. Under QE, on the other hand, there still appears to be a public preference for non-cash assets, or, more precisely, the public does not have the option of eliminating interest-bearing assets from its portfolio, as it would in under an entirely exogenous interest rate policy setting.

If, however, the central bank were to stand ready to buy and sell all government bonds at fixed yields, then there would be complete certainty as to the future rate of interest. It would be tempting to conclude that the theory of liquidity preference would predict that, without uncertainty, the rate of interest would fall to zero. With complete certainty as to future sale prices of bond holdings, there would be no need to hold a money asset. Instead, the rate of interest would be fixed by the central bank, and the public would request money on demand for purchases and then reflux any remaining balances for bonds. In the absence of 'uncertainty about the future rates of interest...the propensity to hoard...will always be zero in equilibrium' (Keynes, 1936, pp. 208-209): a Walrasian general equilibrium. Thus, liquidity preference can determine the rate of interest 'even if "idle money" were zero' (Robertson, 1940, p. 12). 'In the light of Keynes's theory, society faces a *choice* of interest rates' (Chick & Tily, 2014, p. 697, original emphasis).

This portrayal of the central bank as a dealer in an arbitrary variety of bonds should be

connected to the ontological view of money as credit. The ability of the central bank to establish administered prices in a range of assets and interest rates is an important supporting argument in favour of a view that the central bank is in effect purchasing financial assets with its own credit, although the determination of the value of this credit has not been established. A policy of unlimited two-way open-market operations would have an effect on the value of the central bank credit. At the time of writing, the Bank of Japan has a target yield on ten-year Japanese Government Bonds of 0%, and many central banks (European Central Bank, Bank of Japan, Sveriges Riksbank) have negative interest rate policies (NIRP) on commercial bank reserves. The NIRP (see Section 6.3) shows that central banks can affect, not only the interest rate in the interbank market, but also the holding cost of central bank reserves. In Keynes's expression for the total return of assets (see Section 8.2), negative interest rates are the carrying cost associated with reserves, with the implication that, if the carrying cost outweighed the Keynesian abstract liquidity premium, then they would cease to be monetary.

The very existence of the shadow banking system, with its version of liquidity creation, raises questions about the idea that central banks have an unlimited ability to create money. In positing exogenous interest rates, endogenous money theorists appear to be asserting that central banks have no restrictions on money creation. In practice, however, good quality collateral must be exchanged for all central bank money created (Bagehot, 1873); 'it is characteristic of modern systems that the central bank is ready to buy for money at a stipulated rate of discount any quantity of securities of certain approved types' (Keynes, 1930b, p. 211). Unless they are willing to expose themselves to credit risk, central banks do not create money *ex nihilo*—in practice, collateral is required, although the central bank can vary the quality of the collateral. Thus, endogenous money theorists may make too much of the lender of last resort (LOLR) aspect of a central bank's role, banks may have an insufficient amount of good quality collateral, once necessary haircuts are applied:

> When the Bank of England lent to Northern Rock in 2007, it was possible to predict when the LOLR assistance would reach its maximum limit. The limit was duly reached on the date predicted and the government had to take over the financing of the bank and the associated credit risk. (King, 2016, p. 205)

Endogenous money theory requires substantial modification when it is appreciated that central banks require collateral as part of their money creation. The innovation or important

insight is the function of collateral. To understand the importance of collateral, we need to consider the implications of synthetic finance.

The process of bank money creation is an example of a synthetic financial instrument with no endogenous supply constraints (see Section 8.6). The bank and the borrower together create a new synthetic financial asset satisfying the basic property of exchange, namely the 'realization of a *symmetry* between the economic object and its image of value as money' (Lozano, 2015, p. 2, original emphasis). In this case, the economic objects in question are credit and bank money, and the image of value is measured in the unit of account. Synthetic financial exchange can be achieved by any conceivable credit and cash flow structure, with the only restriction that it must satisfy the symmetry of exchange measured in the unit of account. For synthetic financial instruments, it is the '*process* of the exchange itself which constitutes the asset, rather than the exchange constituting the process by which some pre-existing assets are exchanged' (Lozano, 2015, p. 47, original emphasis). The simultaneous creation of a loan and deposit, whereby the deposit itself is money, follows the insight that banking is the creation of synthetic financial instruments, in other words, the exchange of IOUs or a derivative security over central bank money (see Section 4.6).

A rejection of the commodity view of money frees the analysis from the restrictions of classical exchange, whereby a money object must be involved in the immediate settlement or payment of the exchange. The idea that money is a singular commodity-like object through which all exchange must be facilitated is ontologically inaccurate. Instead, the tenure of generic or synthetic financial instruments is not limited to immediate settlement, as it is in classical or barter exchange (see Section 8.6). This loosening of settlement horizon opens the way for credit as an independent phenomenon:

> There is virtually no relationship between the legal transaction in which a buyer agrees to pay a certain amount to a seller, in return for an asset owned by the seller or services to be provided by the seller, and the settlement of the transaction in the form of one or more cash flows. (Black, 1970a, p. 3)

This 'extension of the time-horizon of tenure' (Lozano, 2015, p. 24), made possible by generic and synthetic finance, provides a clear ontological distinction from the Arrow-Debreu model of complete futures contracts with payments settled and cleared at the beginning of time. Once this Walrasian classical exchange restriction is removed, a synthetic universe of future settlement contracts opens, revealing an avenue to the

fundamental analysis of credit and liquidity. When exchange can be achieved by credit, the tenure of settlement spans an entirely new temporal spectrum of contracts, denominated in the unit of account, relating to the expected delivery of payment, not the delivery of the commodity. By removing the 'invariance requirement on tenure' of immediate settlement, entirely new risks appear, since 'the properties of maturity and interest rate now differentiate themselves, and the image of the value of the object is free to grow or shrink over the time-horizon of the exchange' (Lozano, 2015, p. 3). Interest rates, as a truly monetary phenomenon, only appear with generic financial instruments (Lozano, 2015, p. 3 & p. 126).

The evolution of finance is a historical process that aligns with the expanding ontological sets of classical, generic, and synthetic assets. As Jefferis (2017) argues, seemingly successful explanations of one crisis, such as Minsky's 'Financial Instability Hypothesis' (Minsky, ([1986] 2008, p. 194), are merely descriptions of the particular historical episode, but do not explain the historical process that leads to them. Instead, Jefferis argues that the process of constructing and calculating the risk of new financial products creates the illusion of new liquidity. Liquidity crises appear as a reflexive failure of risk calculations because liquidity dynamics are 'immanent and appear through constructive processes associated with the arbitrage and commensuration of financial risks' (Jefferis, 2017, p. 91). The unrecognised and poorly understood expansion of the ontological set to include synthetic finance creates a history of epistemological problems.

The ability of synthetic finance to decompose and redistribute risks in practice stumbles on the issue of collateral. The decoupling of exchange and settlement opens the door to new forms of risk as the image value of money embedded in generic financial instruments acquires the ability change over the maturity period. The introduction of the suite of valuation adjustments (known collectively as XVA) into asset pricing is a development, while at odds with the Arrow-Debreu complete markets paradigm, that attempts to mitigate these new risks and recognise the practical significance of collateral and funding costs (see Section 7.4).[57]

[57] Black (1970a) repeatedly stresses that successful risk separation would require collateral but does not

The importance of collateral can be seen in the conceptual framework used by Mehrling (2012) to understand the fledgling 'market-based credit system'.[58] The framework comprises a Capital Funding Bank, a Global Money Dealer, an Asset Manager and a Derivatives Dealer, as stylised actors in the risk decomposition of an asset in accordance with the principles of synthetic finance. The Asset Manager adopts the risk for the entire system by means of credit and interest rate derivatives contracted with the Capital Funding Bank, intermediated by the Derivatives Dealer. As a 'mirror image' of the Asset Manager, the Capital Funding Bank holds the underlying asset stripped of its risk and raises funds in the short-term money market via the Global Money Dealer, using the de-risked asset as collateral. All participants are secured from counterparty risk, and hence insolvency, by these collateral flows.[59] Any asset-price losses suffered by the Capital Funding Bank, which would otherwise cause funding difficulties as the value of its collateral falls below its borrowings, are filled by the collateral posted by the Asset Manager, via the Derivatives Dealer, to cover the offsetting gains on its derivatives. Collateral is the key ingredient to mitigate any temporary liquidity issues that underlie the transfer and pricing of risk, but permanent losses must be met with money payments. A failure in collateral flows leads to liquidity, but not solvency, problems. Nevertheless, the reliance on collateral in the market-based credit system means that any imperfections can lead to a 'downward liquidity spiral' (Mehrling, 2012, p. 111), where funding problems cause asset fire sales, which lead to more funding problems (see also Brunnermeier & Pedersen, 2009).

The ontological recognition that risk and cash flow are not intrinsic properties of an asset implies that the same conclusion can be made for liquidity. The endogenous availability of money from the central bank implies a view of liquidity abundance, but the post-GFC response to the shadow banking system, with its provision of market-based credit and liquidity, shows that collateral must play a theoretically more significant role in resolving

---

develop the idea.

[58] See also Mehrling et al. (2013) and Pozsar (2014) for similar frameworks.

[59] Taken to the limit, where all credit risk is mitigated by collateral, we could conceive of a state of moneyless pure exchange with immediate settlement and therefore a role for the Walrasian auctioneer. The vision is undermined by the practical difficulties of collateral valuation and haircut determination and the hidden assumption of perfect market liquidity.

the uncertainty introduced by synthetic finance.

9.5 Summary

Post-Keynesian monetary theory explicitly considers the role of the banking system in the operation of an economy in which liquidity and money have a fundamental place. The credit nature of money is recognised and essential to both the theory of the monetary circuit and the theory of endogenous money. The theory of the monetary circuit describes the process of bank money creation and destruction and its essential function in the production and sale of goods and services. The relationship between the financial system and saving by the non-bank public is consistent with the ontological view that money is credit. The circuit theory applies specifically to traditional banking and needs to be updated to explain financialisation and the shadow banking system.

Endogenous money theory extends the circuit theory to include a role for the central bank and an explanation for the determination of the rate of interest. Given a pre-arranged line of credit, credit money can be created at the discretion of the borrower, who has a financial option to draw on a source of bank credit at the interest-rate terms agreed when the line of credit was extended. It follows that bank credit is demand-determined and credit money is created automatically.

The limits to pure-credit monetary expansion are less clear in post-Keynesian monetary theory. An appeal to the central bank's interest-rate policy or the state of trade is an insufficient basis for understanding the potential supply of liquidity. Both the active (rather than passive) nature of the banking system in moving liquidity provision outside traditional bank channels and the importance of collateral in the world of synthetic finance are areas that challenge central ideas of money endogeneity. Rather than rejecting the theory of liquidity preference, it should be seen as extending endogenous money theory into the realm of asset pricing. Overall, the post-Keynesian literature provides important insights into the operation of a credit-money system and its relationship with liquidity that are developed further in the next chapter when we synthesise theoretical elements from both Real and Monetary Analysis into a conceptual framework.

Part V Synthesis

The assessment of Wicksellian and Walrasian general equilibrium models, within the context of the taxonomy of liquidity put forward in this study, has revealed the limited applicability of Real Analysis to the analysis of liquidity. The axiom of monetary neutrality means that neither liquidity nor money has a fundamental place in Real Analysis. Similarly, interest rates in Real Analysis are represented by proxies that only partially aligned with phenomena observed in reality. Instead, liquidity and money are introduced as frictions that delay the realisation of the inevitable general equilibrium. The preordained fixed point of convergence provided by the general equilibrium introduces a further weakness that uncertainty reduces over time, as the effects of financial frictions dissipate and the long-run general equilibrium asserts itself. In actuality, the further the time horizon, the more uncertain is our knowledge of the potential outcomes.

Nevertheless, modern financial theory and practice have recognised and engaged with the ontological realisation that risk is not a fundamental property of an asset; risk can be transformed, separated, and transferred. This separation of risk, cash flow and asset ownership has equally significant implications for the ontology and analysis of liquidity: liquidity also cannot be an intrinsic property of an asset. Nevertheless, without a complete treatment of liquidity, the otherwise-insightful consumption asset-pricing model cannot directly inform the analysis of liquidity.

The investigation of theories associated with Monetary Analysis has revealed two further insights. First, Keynes's theory of liquidity preference warrants a significant reinterpretation in light of the taxonomy of liquidity outlined in this study. This taxonomy has allowed the principles of liquidity preference to be clarified and its place as a theory of the rate of interest restored. Liquidity preference expresses the trade-off between assets with price-protection and those relying on market liquidity, with the rate of interest determined by the uncertainty associated with the proceeds of asset sales in organised markets.

Second, the theories of the monetary circuit and endogenous money engage directly with the operation of credit money and describe essential features of the financial system at particular points during its evolution. Once each is placed in its particular context, generalisations can be made and apparent conflicts between the theories of liquidity

preference, endogenous money and the monetary circuit can be resolved.

In Part V, elements of both Real and Monetary Analysis are combined to construct a coherent and holistic structure around liquidity.

10 A Conceptual Framework for a Theory of Liquidity

10.1 Introduction

The economics and finance literature displays conceptual dissonance in its treatment of liquidity and money. This dissonance warrants an investigation into the relationship between liquidity and money in the literature so that their ontological features can be identified and organised into a conceptual framework. The aim of this research is to develop a deeper understanding of liquidity and money and to provide a framework for evaluating their fundamental causal relationships and tendencies. For this framework to have explanatory power and correspondence with reality, it is necessary for it to be built on a firm ontological platform and to satisfy realist criteria for both assumptions and predictions.

Consequently, this study has established a firm ontological grounding for the analysis of liquidity and money. The rejection of the commodity view of money allows liquidity to occupy a financial dimension unavailable in the restricted realm of Real Analysis. This dimension is increased still further by considering credit money, in accordance with Innes (1913, 1914), as a more general concept than that allowed by the state theory of money. The realisation and acceptance that liquidity and money exist in a dynamic hierarchy of credit (Mehrling, 2013a) based on social relations (Beggs, 2012) are keystones for understanding their relationship and constructing a conceptual framework for a theory of liquidity.

Also important are the ontological insights of generic and synthetic financial exchange (Lozano, 2015), which increase the understanding of risk and its transferability, and also the nature of liquidity itself. The troubled relationship between liquidity and money is evident in the difficulty in resolving the liquidity abundance view of finance and liquidity scarcity view of economics (Mehrling, 2000b). The key characteristic of money is that it can achieve balance-sheet clearing, and the interpretation of the theory of liquidity preference (Keynes, 1936) as an explanation of the trade-off between asset pricing and credit clearing reconciles this apparent tension.

This chapter answers the two research questions posed in Section 1.2. The first section explores and clarifies the relationship between liquidity and money. Then, in the following sections, a conceptual framework for a theory of liquidity is constructed.

10.2 The Relationship Between Liquidity and Money

Answering the first research question requires the clarification of the relationship between liquidity and money in the economics and finance literature and the identification of the flaws in this relationship. At the outset, it can be observed that the relationship is largely determined by the ontological view of money adopted. In many cases, the relationship is dominated by the mistreatment of money, which consequently restricts the number of forms of liquidity that can be represented. In the tradition of Real Analysis, the literature attempts to treat money, even in its fiat form, as a commodity. Money, by this reasoning, is the most liquidity commodity in that it is readily acceptable in exchange for any other commodity. The ontological justification for this treatment is weak, and a recognition of money's credit nature changes its complexion and renders many aspects of monetary theory in Real Analysis defective. As a consequence, liquidity, which is fundamentally defined by its relation to money, tends to be assumed away or is limited by the *ad hoc* nature of the associated money object.

The relationship between liquidity and money can be clarified by the insights of Black (1970a) that risk, being transferable, is not an intrinsic property of an asset and that the timing of cash flow is entirely separate from the moment of exchange. If all the risks associated with an asset can be transferred, the question arises as to where the property of liquidity resides; it cannot simply be an intrinsic property of an asset. Black's insights also point to the existence of generic financial instruments which extend the tenure of exchange beyond the immediate settlement of pure or classical exchange (Lozano, 2015). Generic financial instruments introduce the property of maturity, and the risk of volatility in their value over this new temporal dimension brings with it a monetary rate of interest unrelated to intertemporal consumption preferences. Instead, the rate of interest represents a preference for the timing of cash flows decoupled from the risks and assets used to support the value of these future payments (see Section 9.4).

Unlike pure exchange in neo-Walrasian theory, which is restricted to immediate settlement, generic financial exchange allows payments to take the form of credit promises at any point in the future (or past). The image of value as money that satisfies the symmetry of exchange can take the form of a cash flow promised for any time—past or future. The transfer price of risk established by standard valuation techniques is expressed in reality by a credit flow located in an entirely separate temporal dimension. Even under the assumption that these

credit promises are riskless (that is, expected to be paid with certainty), their extended exchange tenure means that the value of these generic financial instruments can change over their lifespan. This variation in value introduces risks not captured by the concept of a Walrasian auction conducted at the beginning of time. The dimension associated with payment or settlement cannot be represented in a system restricted to pure exchange or perfect barter, which maintains the invariance restriction on tenure associated with classical exchange.[60] Real Analysis, therefore, disregards the real-world process of exchange whereby an asset or commodity is realised, or becomes perfectly liquid, in the form of its image of value as money.

The lack of this temporal dimension, where liquidity is naturally represented, contributes to the deficient treatment of money and the rate of interest in neo-Walrasian general equilibrium models. Neo-Walrasian theory provides a means of determining relative value and, coupled with complete markets, introduces the concept of a risk-free asset with its return based on the ratio of prices of contingent claims. The combination of perfect market liquidity and an asset paying a risk-free rate of interest dominates and obviates any need for a monetary asset. As a result, the neo-Walrasian paradigm encompasses the determination of relative values, but leaves liquidity, money, and interest rates unexplained. Removing the assumption of complete markets only partially undermines the theory since it can be reconstructed with dynamic trading and Rational Expectations. The assumption of perfect market (and funding) liquidity persists.

Instead, the rate of interest is introduced into neo-Walrasian models by coupling the complete markets paradigm with the Wicksellian natural rate of interest, thus tying the risk-free rate to the marginal productivity of capital. The model is extended still further to include the Fisherian theory of intertemporal consumption preferences. In Real Analysis, then, the real risk-free rate of interest is determined by time preference and the marginal productivity of capital. The framework is summarised by the Euler equation (6.6), which determines the representative agent's optimum intertemporal consumption plan, given the assumption of the full employment of resources.

It is evident, however, that the resulting frameworks suffer from deficiencies in their

[60] The Real Bills doctrine can be interpreted as a practical attempt to enforce the restriction of classical exchange necessary for Real Analysis.

interpretation of liquidity and money. Neo-Walrasian theory crowds out any form of liquidity other than perfect market liquidity (Mehrling, 2011) and excludes money altogether (Rogers, 1989, 2014). Paradoxically, the attempt to treat money like a commodity leads to the non-monetary theory of neo-Walrasian general equilibrium founded on pure exchange. Pure exchange, in which all commodities are directly exchangeable for all other commodities, restricts money payment and settlement to an abstraction conducted at the beginning of time. The paradox is that the assumption of perfect market liquidity leaves no role for, not only credit or credit money, but also commodity money.

The Walrasian auctioneer ensures that exchanges of credits denominated in the abstract unit of account represented by the *numéraire* are cleared at the beginning of time. In this way, the theory abstracts from the process of creating and clearing credits conducted by the banking system. Attempts to include money into such liquidity-free models then take *ad hoc* forms, such as: unspecified utility arising from liquidity; search or coordination frictions that cause illiquidity; transactions costs minimised by a spontaneously tradable commodity; or entirely cashless models. Instead of the banking system facilitating and enabling trade in the neo-Walrasian models, 'financial frictions' are introduced that reduce and restrict opportunities for trade. Liquidity morphs into illiquidity frictions (Rogers, 2008a, 2014). Conceptual frameworks in the tradition of Real Analysis (Diamond & Dybvig, 1983; Holmström & Tirole, 1998; Tirole, 2008) represent 'liquidity' as losses due to production inefficiencies and interruptions, unrelated to money. Liquidity and money are represented by concepts that bear little resemblance to, or behavioural similarities with, their real-world counterparts.

Furthermore, the Wicksellian natural rate of interest can only find theoretical justification in a world with a single commodity (Rogers, 1989). Imbalances between saving, investing and consuming a single commodity, say corn, can be rectified simply by consuming or planting. Perfect technological liquidity crowds out any other form of liquidity. Credit money does enter Wicksellian theory, but only to ensure price determinacy—an unnecessary safeguard when the solitary consumption good already acts as *numéraire*. The analysis of Wicksellian theory has shown that the productivity of the system can be aligned with any exogenous interest rate and should be abandoned as part of the theory of interest (see Section 6.2). Finally, in Fisherian interest rate theory and neo-Walrasian general

equilibrium, the supposed intertemporal characteristics introduced by time-dated commodities are no different from analysing commodities in different locations, so that the 'interest rates' derived from the model are no more than the relative propensities to consume in different locations. The Wicksellian and Fisherian concepts only have validity in a single-commodity world. Furthermore, in a model with more than one consumption good or incomplete markets the anchor of the risk-free rate is lost. Without the risk-free rate, modern asset-pricing theory has no anchor, and the number of valid discount factors becomes infinite (see Section 6.6).

The recognition that risk and cash flow are ontologically separate from assets is reflected in the widespread adoption of the complete markets paradigm for asset pricing. Asset pricing, based on the valuation of risks and the price of risk transfer, is consistent with this ontological principle, but is also constructed on the basis that all such risks have perfect market liquidity at their fundamental value. It is a paradigm without a place for liquidity. To understand liquidity, we must move beyond Real Analysis and introduce generic financial assets, which loosen the restriction of classical exchange and the immediate settlement inherent in pure-exchange models. Once the invariance requirement of tenure is removed, then cash flows and payments are decoupled from exchange and interest rates and credit risks appear naturally. Similarly, the role of a purely liquidity-derived discount factor, which expresses the liquidity premium derived from the theory of liquidity preference, becomes apparent. The fact that cash flows can be separated from assets and vary according to an entirely new temporal dimension introduces the need for the rate of interest to equalise the demand and supply for generic financial assets at each point on the maturity spectrum, to balance tenure preference.

The behavioural factors behind these demand and supply curves, however, are those expressed in the theory of liquidity preference (Keynes, 1936), not loanable funds (Robertson, 1940). Money is credit, and the rate of interest is the cost of delaying settlement. With a credit view of money, it follows that the choice between holding money or bonds is not whether to invest or not, as it is in the loanable funds theory. Instead, it is the decision on how, or equivalently in what form, to invest. The social relationship underpinning money negates the idea that hoarding has the direct effect of rendering money unavailable for investment. The loanable funds theory, whereby a fixed stock of funds is either hoarded or available for investment, is a fallacy based on the erroneous commodity

view of money.

The theory of liquidity preference fills this lacuna and permits the pure rate of interest to be determined in a dimension entirely missing from general equilibrium theory, which contends with value, but not cash flow or liquidity. Value and cash flow can be entirely separated, and this ontological observation points to a path to a reconciliation with general equilibrium theory, which limits itself to the utility of commodity exchange. By removing the QTM as an explanation of prices, and marginal productivity as an explanation of the rate of interest, general equilibrium models can be extended to include exchange tenure and hence liquidity. The stochastic discount factor must be derived from liquidity preferences, and then equilibrated with the intertemporal rate of substitution.

Similarly, the theory of endogenous money (Lavoie, 1984, Moore, 1988) explains why the banking system is not physically constrained in its provision of money in the way necessary to restrict lending to the supply of loanable funds. In addition, the theory of the monetary circuit further illuminates the relationship between liquidity and money within the circular flow of income. Nevertheless, the banking system must manage its exposure to tenure preference; the liquidity mismatch generated in the banking system by the tenure preference of borrowers and lenders must be resolved by the complex of interest rates prevailing in the money market. It is not that interest rates equilibrate the supply of and demand for loanable funds, it is that they equilibrate borrowers' and lenders' preferences for length of exchange tenure.

An explanation of and a distinction between liquid and illiquid forms of assets are missing from Real Analysis. A more promising avenue follows an investigation of the effect of liquidity on time preference. Also essential is an evaluation and selection of the aspects from the tradition of Real Analysis that make a valid contribution. We turn next to the construction of a conceptual framework to increase our understanding of liquidity.

10.3 A Conceptual Framework

The second research question asks whether it is possible to construct a conceptual framework for a theory of liquidity that can guide future research. The recognition that the tenure of exchange can expand to allow for more than immediate settlement transforms the understanding of both liquidity and money. Delayed settlement, which manifests itself in the form of generic financial assets, introduces credit at a fundamental level. Money appears

in many forms as part of a dynamic credit hierarchy and its social-relational aspect comes to the fore. The result is a structure and dimension of payments vastly beyond that embraced by models based on pure exchange and commodity utility that take the focus in Real Analysis.

Once it is recognised that money is a social relation that always has a debt counterpart and is exposed to the potential operation of the reflux mechanism (Lavoie, 1999), it follows logically that holding money is an asset allocation decision, and that Keynes's two-asset model is a valuable pedagogical device for representing the resulting liquidity dichotomy. This dichotomy, which bears superficial resemblance to the distinctions between inside and outside liquidity, and between public and private liquidity, can be visualised as two forms of financial put option, categorised according to the nature of the liquidity supplied by dealers or financial put option writers. Inside/outside liquidity and public/private liquidity, as discussed in Section 5.6, inherit too much from their association with inside and outside money, with their undue emphasis on state money. The reduction of emphasis and placement of state money in a hierarchy of credit allows for a more far-reaching study of liquidity. Instead, this liquidity dichotomy takes central place in the conceptual framework presented here as a distinction between exchange liquidity and redemption liquidity.

The dichotomy between exchange and redemption liquidity builds on the ontological insights of generic financial exchange (Lozano, 2015). The property of maturity and extended tenure in exchange, allows monetary settlement to be entirely independent from contractual and physical exchange. This dimension, missing from much of the literature on liquidity and money, is governed by the behavioural characteristics underlying tenure preference, and allows for a realistic introduction of the rate of interest. The basic property of tenure in generic exchange undermines theories of the rate of interest in the tradition of Real Analysis. Cash flows are independent from the time of exchange and, therefore, consumption itself, and so the monetary rate of interest cannot represent a Fisherian trade-off between consumption in the present and the future.

More important, liquidity is the product of this structure and dimension of payments and the dynamic interaction with the tenure of exchange. The loosening of the tenure of exchange, so that payment for a commodity exchanged today is decoupled from the moment of exchange and is pushed into an uncertain future, allows for the independence of cash flows from the utility of consumption and the productivity of production.

Consumption today can be supported by commitments to payments in the future, at the maturity of a generic financial asset that satisfies the required image of the commodity's value in the unit of account. This generical financial asset, which only provides guaranteed liquid value at maturity, imposes a liquidity restriction on its owner. The owner, in a desire to break this contractual tenure, must appeal to the asset's exchange value, and this value is vulnerable to the conditions prevailing at the moment of sale. The acceptable tenure of exchange—longer for buyers and shorter for sellers—depends on the balance of preferences for guaranteed redemption liquidity versus uncertain exchange liquidity. These are behavioural drivers independent of intertemporal consumption plans or production possibility frontiers; they are based on the supply and demand for liquidity.

Therefore, the liquidity dichotomy, coupled with the ontological insights arising from generic finance, leads to an understanding of the relationship between, and a possible reconciliation of, the neo-Walrasian theory of value with the liquidity preference theory of the rate of interest. The neo-Walrasian theory of value, for which a theory of the rate of interest is absent, is specified within a framework where the tenure of exchange is immediate. Loosening this restriction on the tenure of exchange brings forth a temporal dimension that forms the basis of a theory of the rate of interest. Coupled with the theory of physical exchange provided by the neo-Walrasian framework, the theory of the rate of interest based on the tenure of exchange provides a valuable avenue to the reconciliation of aspects of Real Analysis and Monetary Analysis. The long-period equilibrium specified by a neo-Walrasian commodity exchange model should be anchored to the rate of interest specified by the tenure preferences of borrowers, lenders and the banking system.

Keynes's theory of liquidity preference provides the point of departure for a conceptual framework organising these characteristics of liquidity. With the distinction between redemption liquidity and exchange liquidity made explicit, the theory of liquidity preference and Keynes's two-asset model are sufficient to present a theory of the rate of interest in the tradition of Monetary Analysis whereby the uncertainty regarding the future rate of interest determines the current rate of interest.

10.4 Exchange Liquidity and Redemption Liquidity

Exchange liquidity and redemption liquidity can both be conceptualised as a financial put option provided by a liquidity provider. Just as risk is not the specific property of an asset,

it follows that exchange liquidity and redemption liquidity are not intrinsic properties, and are simply applied by a potentially unrelated third party. The difference between exchange and redemption liquidity stems from the nature of the liquidity put option provided to its holder. The crucial distinction is whether the put option is over the option-writer's own liabilities or the liabilities of others.[61] Redemption liquidity is provided by agents over their own liabilities. Money issuers offer redemption liquidity by allowing for immediate debt clearing and reflux.

Exchange liquidity in an organised market is a financial put option provided by a market maker (see Section 5.2). The put option manifests itself through an entitlement to sell or buy at the 'exercise price' stipulated by the market maker's respective bid or ask prices. The option premium embedded in the bid-ask spread compensates the dealer for the provision of the liquidity option. The provider of exchange liquidity has no direct relationship with the asset, but merely facilitates its shiftability, confirming that exchange liquidity is not an intrinsic property of an asset. Exchange liquidity encompasses both market and funding liquidity and is microeconomic in nature. Exchange liquidity is an unreliable mirage that disappears in a crisis.

Dealers provide exchange liquidity, which is the mechanism by which a pre-agreed tenure of exchange can be reduced for the generic-asset holder. By providing exchange services for generic assets, dealers mitigate the need to wait until the asset's maturity for it to become liquid in the final settlement of the generic exchange. The dealer can buy the asset outright or provide funding liquidity in the form of a collateralised loan, such as repo. In this way, dealers offer the ability for an agent to adjust the tenure of generic financial assets. Consequently, dealers operate outside the realm, and hence are independent, of any hypothetical long-period equilibrium based on pure exchange. Their role is to equilibrate the supply of and demand for the length of tenure in generic exchange.

Exchange liquidity encompasses the idea of liquidity when defined as the conversion into money within the three dimensions of certainty, short notice and without loss (Bronfenbrenner, 1945; Hicks, 1962; see Section 8.2). By invoking an idea of loss, it builds

[61] In terms of generic financial instruments, one person's liability is another person's asset. In the case of physical assets without technological liquidity (unlike corn, which is technologically liquid, see Section 5.1), only exchange liquidity can apply.

on a theoretical platform of fundamental or long-period prices provided by a general equilibrium model (see Section 6.5). In practice, however, exchange liquidity relies on the existence of market makers and money dealers, who are prepared to provide shiftability based on relative value. Dealers do not attempt to measure fundamental value, nor do their profits rely on assessing the difference between the market price and fundamental value. Instead, they attempt to operate a 'matched book' so that systematic risk is reduced as far as possible.

It is precisely because exchange liquidity cannot be an intrinsic property of an asset, that analysing liquidity using the three dimensions of certainty, short notice and without loss, which rely on the existence of fundamental value, is paradoxical and inappropriate. The notion of fundamental value, derived in a general equilibrium assuming perfect exchange liquidity, falls into the trap of asset fetishism (Lozano, 2015), by ignoring the ontological principle that risk and cash flows—and not assets *per se*—are the economic phenomena that can be valued (see Section 8.6). The concept of exchange liquidity cannot be used to understand the liquidity of cash flows, which, by their very nature, are self-liquidating. Neo-Walrasian general equilibrium models lack financial instruments that can represent the image of value in exchange by denying the distinction between value and cash flow. The liquidity of value and the liquidity of cash flow must be considered separately.

A distinction must be made, therefore, between the variability of expected returns across short- and long-dated assets, and any liquidity premium required for an asset supplied with a dealer put option. Short-dated assets are more liquid in the sense that their closer maturity date entails an earlier repayment or liquidation date, which a long-dated asset does not (see Section 5.5). Longer-maturity assets rely on exchange liquidity provided by a dealer in an organised market to be immediately realised in money terms. Exchange liquidity is fundamentally uncertain and corresponds with the empirical observations of time-varying discount factors (Cochrane, 1992; Shiller, 1981; see Section 6.6), and agents attempt to counter this uncertainty with judicious allocations of assets with price-protection and redemption liquidity.

To repeat, exchange liquidity, being a form of liquidity supplied primarily by dealers, is not an intrinsic property of an asset, since its purpose is to alter the original terms of the exchange tenure. This being the case, exchange liquidity is ordinal only via a complex functional relationship (see Section 8.2). Any ranking of assets in terms of exchange

liquidity would be conditional on the dealer structure prevailing at the time of measurement and would therefore only be transitory. Without a fundamental value for comparison with a realised value, transactions cost methods also fail as ordinal measures.

The liquidity of money arises from its ability to achieve immediate balance-sheet clearing between the issuer of the money and the holder of the money (see Section 4.8). Money, being a synthetic financial instrument arising through the exchange of debts, enables credit clearing and provides the ready means to dissolve its associated social relationship. Money provides reflux to its issuer, an action by which the money holder avoids the need to settle their debt to the issuer with goods or services. The redemption liquidity associated with money contrasts with the exchange liquidity facilitated by a market maker. Money's liquidity, or immediate reflux to the issuer, is an essential feature. Like exchange liquidity, however, redemption liquidity is not an intrinsic property since it must be provided by the money issuer.

Importantly, the difference between money and non-monetary debt can be delineated by this reasoning: only a credit that can be used to repay an issuer immediately can be money. If the issuer is a large economic agent, to which many other agents are constantly indebted, like the state or the banking system, then these credits will be in continuous demand as a means of debt repayment. Although state fiat money appears to be irredeemable, it is sufficient that it can be used to settle debts arising from taxation for it to have value (see Section 3.8). Similarly, a bank loan cannot be repaid without first securing possession of an offsetting deposit with that same bank (see Section 4.2). Both forms of money offer redemption liquidity. Thus, the crucial difference between credit money and shadow-bank 'money' or repo is that the liquidity of repo is founded on the market liquidity of its underlying collateral and is therefore reliant on exchange liquidity. Given the microeconomic nature of exchange liquidity, the failure of shadow-bank 'money' during the systemic GFC should be unsurprising.

Money is a credit that can be used to repay a debt to the issuer of the credit. It is a synthetic financial instrument without maturity, or, equivalently, a perpetual asset with a redemption put option written by its issuer (see Section 4.8). If the credit is dated then it cannot be money since its ability for use as repayment has been postponed—it offers no redemption liquidity. Demand deposits, on the other hand, are money because they can be used to extinguish a loan obligation to a bank. Government bonds are not money because they must

be converted, by means of exchange liquidity, into central bank reserves or state money to be used to pay taxation. Central bank reserves, even when they pay interest, differ from bonds, which do not have redemption liquidity. The distinction between exchange and redemption liquidity allows for a deeper understanding of the supply of liquidity and money.

10.5 The Supply of Liquidity

Immediately redeemable credits are money and the value of these redeemable credits relies on the balance between amount issued and the availability of offsetting credits. If more credits are issued than can be supported by available offsets, then their value will be reduced to zero. In the case of state money, this reduction would be due to an imbalance between money and taxation. Unless offsetting credits can be sourced, a debtor (or credit issuer) must redress any credit-debit imbalance by reducing the amount of their debts immediately redeemable. In other words, they will need to exchange their redeemable credits for new generic financial assets with the property of maturity. This action is a form of own-funding liquidity, whereby own-issued money is sourced. Own-funding liquidity, achieved by selling a new generic financial instrument, is then subject to the cost of postponing immediate reflux.

The maturity of a financial instrument able to serve as credit money is crucial: debts can only be repaid with immediately redeemable or undated credit (Innes, 1913, 1914; Tymoigne, 2017; see Section 4.8). A debtor can repay a creditor with an offsetting credit against that same creditor, but only if the offsetting credit is currently due for repayment. The creditor can reject as payment their own debts if those debts themselves are not due for payment. In this way, creditors can control the reflux mechanism by managing the maturity profile of their debts. Pushing the repayment dates for debts into the future is an effective method of maintaining the balance of debits and credits currently due. Thus, the property of tenure embedded in generic and synthetic financial assets underpins the concept of redemption liquidity.

The condition to be satisfied for the postponement of payment, or lengthening of maturity, is the rate of interest. For example, government bonds cannot be ‘spent’ (see Section 6.4) because the government has purposely converted its immediate, monetary debts into those not due until a future time and is under no obligation to accept these future debts as payment

for current taxes. The value of the government's immediately redeemable credits, in the form of state money, is based on its current period spending and tax flows. Current spending not matched by tax flows must be postponed if the value of state money is to be maintained.

Money can only be refluxed back to its issuer, and the issuer determines how much of its money is available for reflux by liability management through interest rates, thereby maintaining its value to other agents for use as a medium of exchange. For an issuer of debt, postponement is the alternative to a reduction in the value of immediately payable debts. The amount of immediately payable debt must be altered by extending it into the future, but at a cost. This cost, expressed as the rate of interest, provides the inducement for some agents to postpone their attempts at reflux. Thus, the issuing of debt, or accessing own-funding liquidity, is itself a form of reflux, necessary to prevent the value of money falling to zero, in other words, the issuer defaulting.

Again, this process is seen in the management of government debt, where swapping state money for non-monetary interest-bearing liabilities removes the need to alter the level of taxation. Similarly, banks, via their liability management, carefully maintain an asset-liability term profile so that longer-term assets are funded by longer-term liabilities. Banks cannot create unlimited amounts of monetary liabilities; they require longer-term funding which cannot be used for repaying bank debts or conversion into state money. Banks pay interest on deposits that cannot be redeemed to prevent the reflux of their liabilities and thereby maintain their value at par with state money.

Money issuers must be cognisant of their place in the hierarchy of money, which is constructed explicitly in terms of a gold standard or commercial bank convertibility to state money, or implicitly, in terms of an inflation or exchange-rate target. When the nature of the liquidity put option offered by money issuers depends on convertibility to higher forms of money, then the resulting hierarchy of money signifies that, unlike exchange liquidity, redemption liquidity is ordinal. The lever for managing the value of money is the interest rate on the non-money debt that the issuer offers to discourage reflux or, equivalently, to prevent the liquidity put option from being exercised. Money issuers who aim at a target real value of their money must offer a reflux exit point so that money holders can adjust their desired stock of money. The relevant insight expressed in the endogenous money literature is that, if a money issuer sets the terms under which money-debt can be exchanged for non-money debt, then the balance between money and debt will be determined, not by

the issuer, but by the holders.

Similarly, the government, as one of the issuers in the hierarchy of money, must postpone the process of reflux for its monetary liabilities to maintain price stability. This mechanism should not be confused with a quantity theoretic explanation of prices. A monetary balance must be maintained primarily against the level of taxation, and not between the stock of money and the productive capacity of the economy, as per the quantity theory of money. At issue is not the size of government debt, or the monetary base, but the liquidity mismatch between taxation and monetary liabilities. These are either resolved by conversion of money into debts or a discount on government liabilities.

Historically, monetary assets have taken different values even when measured in a single unit of account (see Section 4.3). Against competing monetary assets, state money has not always played the role of the strongest money. In general, the banking system maintains a peg to the government's currency thereby fixing the value of bank money in the unit of account. Otherwise, as has happened in historical periods, the banking system may be able to command a premium over the value of the government's currency, depending on its integration and its liquidity management. The value of money is based, not on the total amount issued in the overall debt sense, but in the balance between purchasing power and credits owed in the current period, or the ability to support the reflux put option.

Hence, the stock of money comprises immediately payable, or undated, debt, which must be repaid on demand, and so can itself be used for repayment. Conceptually, this stock is not in continuous existence throughout time; it is either repaid, and thereby destroyed, or rolled forward. Rolling a debt forward is fundamentally equivalent to repaying and simultaneously re-borrowing. Conceptually, there is no externally set amount of money that forever circulates with a velocity, but, instead, a continuous flow of creations and cancellations—efflux and reflux—whereby each obligor's creations and cancellations must balance if their credits are to maintain their value in terms of the unit of account. The theory of the monetary circuit applies to all money issuers.

Nevertheless, credit money emitted by an agent will circulate before it refluxes, either as repayment or extended by postponement. A decrease in tenure preference will result in an increase in the demand for credit money and its circulation time will increase as credits are hoarded. As a result, some agents, who are indebted to the credit-money issuer, will struggle to obtain the credits necessary for repayment, and will need to borrow them from others.

This scramble for credits will increase the interest rate on the inter-agent borrowing and lending of these credits. Similarly, the cost to the issuer of postponing reflux, or own-funding liquidity, will increase. In response, the issuer can alleviate the shortage by purchasing the generic financial assets of others with newly created credit money, in other words, by conducting open-market operations or QE.

The central bank's convention of using open-market operations as its monetary mechanism applies a liquidity option to those assets that the central bank receives as collateral, mostly government debt. Again, this viewpoint shows us that, government debt is not redeemably liquid. Instead, government debt gains exchange liquidity from the central bank's dealing in the government debt market. The public-private form of liquidity, where government bonds underlie private repo, is an extension of this central-bank provided exchange liquidity (Pozsar, 2014, 2015; see Section 5.7).

Price-protected instruments, which provide price stability in the unit of account (Ricks, 2011), are not simply limited to the forms of outside money. Bank deposits, as a medium of exchange, provide price-protection and redemption liquidity by means of reflux and loan repayment. Less successfully, the shadow banking system attempted to create redeemably liquid instruments. The shadow banking system combined market liquidity and funding liquidity to create price-protected assets, in an attempt to increase the supply of redemption liquidity. The innovation of increasing the liquidity supply by means of security repurchase agreements relied on the ability of exchange liquidity to simulate the redemption liquidity of a price-protected asset. The ability to synthesise redemption liquidity as shadow-bank 'money' (Gabor & Vestergaard, 2016) was tested and subsequently failed during the GFC.

Overall, the supply of liquidity is offered by dealers that issue liquidity put options, and these issuers form a hierarchy. In order of strength, they are generally the government, the central bank, commercial banks and securities dealers, although this order is not immutable. The hierarchy aligns with the public-private, outside-inside spectrum of liquidity identified in Chapter 5, and straddles the exchange/redemption dichotomy. The value of each of these put options depends on the perception of the default risk of the issuer, determined by the balance of credits payable and receivable. For instance, the exchange liquidity offered by a security dealer is unreliable, since it is implicit and can be withdrawn during periods of market turbulence. Nonetheless, all liquidity is provided by a dealer that stands ready to convert assets, and this set of dealers should include traditional money issuers, such as

central banks and commercial banks.

The clarification of the supply of liquidity allows us to conduct a similar exercise for liquidity demand.

10.6 The Demand for Liquidity

Generic finance complements the view that money is credit with an immediate redemption option, situated on a credit spectrum of delayed settlement that expands the tenure of exchange. This spectrum is independent of any particular asset, since risks and cash flows can be separated from the assets that they reference. Money itself is a synthetic financial instrument that preserves the invariance requirement of immediate settlement associated with classical exchange, thereby eliminating the risk associated with a decoupled time horizon. In models with a single fixed moment of exchange, such as that associated with a Walrasian auctioneer or the portfolio rebalancing model of Tobin (1969), the concept of a risk-free asset is sufficient to function as the determinant of the rate of interest, since the invariance requirement of immediate settlement is strictly enforced. Without these restrictions on exchange settlement, the rate of interest is instead required to equilibrate the supply and demand for generic financial instruments.

This dimension of credit and liquidity—and the rate of interest that arises with it—is unrelated to assets *per se*. The cash flows associated with any particular asset can be de-coupled from it, and the creation of generic and synthetic instruments, which separate risk from assets, confirms that liquidity is also not a property of any specific asset; it is a property of the nature and the maturity of cash flows. It is informed by the knowledge that the value of generic financial instruments, although perhaps conceptually in equilibrium at any point in time, can change over their life. The extra time dimension related to generic and synthetic finance creates entirely new risks missing from Real Analysis.

It follows that, with the credit nature of money in its private and public forms, it is not possible to simply refer to ‘money’ as a single concept. More generality can be introduced by considering the taxonomy of liquidity represented in Table 2. It then remains to consider the nature of the demand for liquidity. By clarifying the different forms of liquidity, this study has shown that that the theory of liquidity preference is founded on the interplay between assets with price-protection or redemption liquidity, and those with only market, funding or exchange liquidity. It was established in Chapter 8 that market liquidity and

funding liquidity, provided by organised markets, are important pre-conditions for the speculative motive, but do not represent the essential form of liquidity underlying Keynes's theory of liquidity preference itself.

For example, liquidity preference is not satisfied by assets, such as long-term Treasury bonds, that trade in, what are often referred to as, deep and liquid markets and can be realised for cash as short notice. Although such an asset can be realised for an amount of cash that is (presumably) close to its 'fair value', the actual amount of cash is unpredictable and is therefore not suitable for meeting an unexpected cash payment of possibly knowable size. It is for more than pedagogical reasons that Keynes's presentation of liquidity preference places a sharp distinction between liquid assets and non-liquid assets, in a model comprising only cash and bonds. Liquidity preference is not founded on the market liquidity or funding liquidity displayed by bonds; it expresses the desire for assets with price-protection, information insensitivity, or redemption liquidity. Liquidity preference motivates investors to hold assets that maintain their value against the unit of account, instead of exchange liquidity.

As identified by Bibow (1998), in Keynes's framework of expected returns, any premium associated with exchange liquidity should be interpreted as a holding cost (see Section 8.2), with the consequence that the large empirical literature on bid-ask spreads is not directly relevant to the theory of liquidity preference. Analysis of the dealer microstructure provides insights into the profit-making activities and asset inventory management of dealers but does not contribute to a macroeconomic theory of liquidity. Ultimately, the microstructure analysis can only reveal the various ways that the carrying cost is expressed: either as a profit to market makers to provide an exchange-liquidity put option; or as the cost of search and delay in transactions. Liquidity frameworks in which illiquid assets have discounted prices do not align with the interpretation of Keynes's liquidity preference framework presented in this study.

Instead, liquidity is realised as an abstract premium earned by the issuer of assets with redemption liquidity, as *seigniorage*. Money, being the liquid asset *par excellence*, is awarded a liquidity premium in Keynes's framework, not illiquid assets. In reality, however, money earns no premium. The fee from lending money arises from not actually holding money, it arises from holding another type of asset entirely: a loan for which the rate of interest is its return. The failure to distinguish the rate of interest from the liquidity

premium of money is a subtle, but significant, error, which has led the interpretation of liquidity preference astray. The liquidity premium that Keynes assigns to money is due to the utility gained by holding it; the premium is not directly apparent in its price. The fact that this utility is observed through the expected return of other assets is an implication of the theory but not the core of the theory itself. The utility attached to the liquidity of money makes it costly to borrow from other agents, and interest rates are evidence of this utility but not its cause. In equilibrium, the rate of interest earned from lending money must equal the liquidity premium on money, but they are not the same thing.

Thus, the essential element at the core of liquidity preference is not the explicit cost of illiquidity, but the hidden premium associated with liquidity. Consequently, liquidity preference cannot be stated in terms of market liquidity or risk preferences as in the models of Hicks (1989) and Tobin (1958). The liquidity premium differs fundamentally from a risk premium by adding an overlay of uncertainty to the assessment of risk, in the form of the Keynesian weight of probability (see Section 8.4). Price-protected assets reduce this uncertainty, since the unit of account is a mapping from all events, of any nature, to a measurable outcome. All events, known or otherwise, are covered. When all outcomes are measured in the unit of account, the range of alternatives is known. The uncertainty associated with unexpected alternatives or possibilities is eliminated; no new outcomes are possible. On this restricted range of outcomes, the estimation of probabilities and their weights is possible, as new information always increases the weight of probability. An asset's return is comparable, the domain of outcomes is manageable, and fundamental uncertainty is partially mitigated.

The price-protection feature, in a probabilistic framework, is the property that an asset maintains a constant price across all future states and sample paths of nature. The desire for liquidity is to avoid uncertainty in future returns and the time-variation in discount factors. The phenomenon behind liquidity preference is precisely the unpredictability of the discount factor and asset-price risk. Redemption liquidity provides a stable value in all outcomes, especially in bad states of nature, where the value of other assets has fallen.

A price-protected asset’s correlation with systematic risk or stochastic discount factor is therefore close to zero. Price-protected assets have a payoff that displays zero correlation with systematic risk and a low risk of default, and this information insensitivity increases their predictability and, hence, liquidity. According to the empirical finance literature,

excess expected returns—those above the risk-free rate—are based on correlation with systematic risk. The zero-correlation feature is commonly associated with the concept of the risk-free rate, which, unlike the rate generally earned on money-proper, is non-zero. The risk-free rate, however, is only appropriate for a well-defined investment horizon, without path-dependency. For example, selling a risk-free bond before its maturity relies on exchange liquidity, which exposes the owner to price risk. In the complete-markets model that forms the theoretical foundation for the risk-free rate, exchange liquidity poses no problem; all assets are tradable at their fundamental values. Instead, as we shall see next, it is the interplay between exchange liquidity and redemption liquidity that determines the rate of interest.

10.7 Liquidity and the Rate of Interest

The following simple demonstration model, an adaptation of the certainty-equivalent model presented by Pratt (1964) (see Section 7.5), illustrates the key concepts of the conceptual framework and clarifies this study's interpretation of Keynes's theory of liquidity preference. Under consideration is the determination of an individual's 'safe' rate of interest (see Section 8.1). The model presented here is highly simplified and is offered as a pedagogical illustration of the mechanism that relates the liquidity dichotomy to the rate of interest. More elaborate constructions are possible, but would merely obscure the main purpose, which is to compare a price-protected asset with one exposed to market liquidity and show how the concept of a safe rate of interest arises. The purpose of the model is to show that it is possible for current interest rates to be determined by uncertainty concerning future interest rates, as opposed to the usual idea that forward rates are predictors of future spot rates.

The entry point of the model is the closing stage of the monetary circuit. Production and consumption decisions have been made, and the agent has wealth W available to invest in an interest- and non-interest-bearing asset. Following Keynes's two-asset model, the agent's investment choice consists of combinations of the money asset, with redemption liquidity, that yields a value of unity at any point over the investment horizon, and a zero-

coupon bond paying an amount $1 + \lambda > 1$ at time T.[62] The bond is tradable at any time $0 < t < T$ by means of exchange liquidity, but only at an uncertain price p_t.[63]

The agent faces the risk that, in relying on exchange liquidity, a loss is incurred at the time of sale, in comparison with an investment in the price-protected instrument. Because expected returns vary, more than can be explained by intertemporal preferences and dividend volatility, the agent faces a path-dependent market-price risk (see Section 6.6). The uncertain dynamics of the bond price impact the agent because, at some uniformly random time $\tau \sim U(0, T)$ between 0 and T, the agent will need to realise their entire bond and money holdings. The agent's liquidity preference therefore reflects the relative utility of exchange versus redemption liquidity.

Even though the proceeds of the bond are uncertain between time 0 and time T, the agent is nevertheless certain of the proceeds from either investment at the investment horizon T. This certainty at the investment horizon, but uncertainty in between, differentiates this model from those of Hicks (1989) and Tobin (1958). This uncertainty determines the rate of interest, unlike in the models of Hicks and Tobin, where risky and risk-free asset returns are given, and the agent's task is to construct an optimal portfolio for a fixed time horizon. In the current model, both assets are risk-free at the investment horizon, but only one is money in that it offers a price-protection feature. Rather than being a risk-free rate, λ is the liquidity premium necessary to compensation the agent for the uncertainty regarding the future rate of interest, in the form of the bond price at the time of sale, as per the precepts of liquidity preference (see Section 8.1). Furthermore, both assets can be classified as near-money in the sense that each is self-liquidating (see Section 5.5), so that the agent's choice is isolated to be between exchange liquidity and redemption liquidity.

The price uncertainty facing the agent is modelled by means of a Brownian bridge process, whereby the bond price p_t begins at $p_0 = 1$ and then fluctuates randomly before finishing

---

[62] The bond pays more than its notional amount at maturity. This deviation from standard bond practice is for simplicity and can easily be corrected by adjusting the initial bond price by the final payment.

[63] Note that λ should be a function of T, since there will be a premium associated with every time horizon which contributes to the complex of interest rates. For simplicity the dependence of λ on T is omitted.

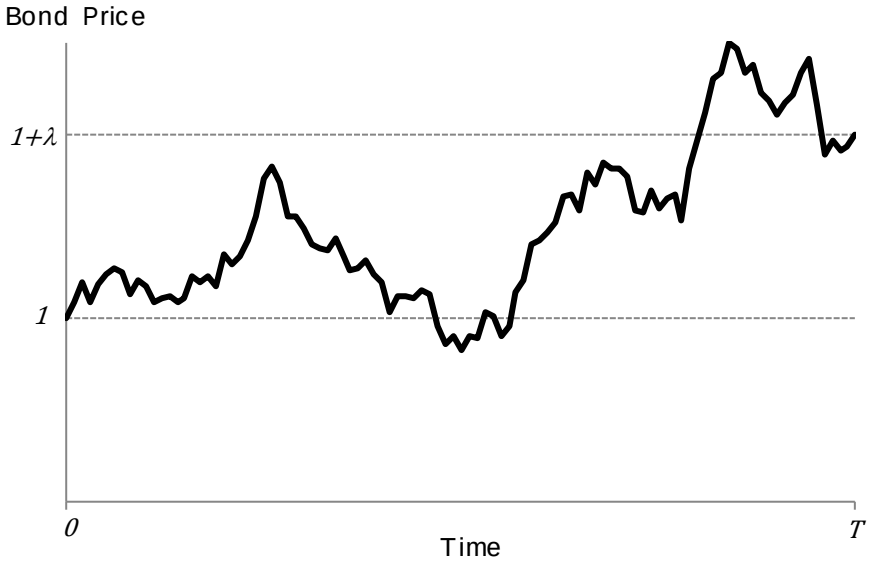
with certainty at $p_T = 1 + \lambda$ (see Figure 1).[64] A Brownian bridge process is used to simulate the capital-uncertainty faced when bonds are not held to maturity (see Section 5.2).

By the properties of a Brownian bridge, p_t is then defined as

$$p_t \equiv 1 + \frac{t}{T}\lambda + B_t - \frac{t}{T}B_T, \tag{10.1}$$

where B_t is a Brownian motion process such that $B_t \sim N(0, \sigma^2 t)$, and $\sigma^2 t$ is the agent's subjective measure of risk regarding p_t. The agent does not know the 'true' dynamics of p_t, but uses the Brownian bridge model to frame the economic effects of uncertainty. By comparison, the money asset maintains a constant value of unity.

Figure 1 – Example of Brownian Bridge Bond Price Dynamics



The agent is risk averse with a constant relative risk aversion (CRRA) utility of wealth W given by

$$u(W) \equiv \frac{W^{1-\gamma} - 1}{1 - \gamma}, \tag{10.2}$$

where $\gamma \neq 1$ is the coefficient of relative risk aversion.

[64] A Brownian bridge is a standard mathematical process that represents a 'pinned' Brownian motion or 'random walk', in which both the origin and destination are fixed but the dynamics in between exhibit a random walk. For more details, see Karatzas and Shreve (1988) or Klebaner (1998).

Following Pratt (1964), the uncertainty regarding the proceeds of the bond implies that the agent will demand a return premium π to hold the bond. This premium π depends on wealth W and the probability distribution of the bond price, and so can be written as $\pi(W, p_t)$. The premium π therefore reflects the safe rate of interest that leaves the agent indifferent between the price-protection of money and the capital-uncertainty of the bond. It is compensation for having to rely on exchange liquidity to meet uncertain payments.

To find the premium π, we substitute p_τ for R and 1 for $\mathbb{E}[R]$ in equation (7.1). We also adjust the size of the necessary premium by the expected time until sale. It follows then that, for the agent to be indifferent between the bond and money, the premium π must satisfy the equality:

$$\mathbb{E}[u(Wp_\tau)] = \mathbb{E}\left[u\left(W(1 - \frac{\tau}{T}\pi)\right)\right]. \tag{10.3}$$

That is, the expected utility from either asset is the same, where the expectation $\mathbb{E}[.]$ is taken at time $t = 0$ over both the random sale time τ and the bond price p_τ at the time of sale, also random.[65] The premium π is scaled by the ratio of the random time of sale τ and the maturity of the bond T.

The fact that $\mathbb{E}[Wp_t] = W\left(1 + \frac{t}{T}\lambda\right)$, and not W, is inconvenient, but can be remedied by setting $p_t^* \equiv p_t - \frac{t}{T}\lambda$ and using the adjustment to the premium π outlined by Pratt (1964, p. 124, Equation 2), repeated here in a more convenient form as

$$\pi(W, p_t) = \pi\left(W - \frac{t}{T}\lambda, p_t + \frac{t}{T}\lambda\right). \tag{10.4}$$

The process p_t^* is a Brownian bridge anchored to unity at both $t = 0$ and $t = T$. Using p_t^* in place of p_t, equation (10.3) then becomes

$$\mathbb{E}[u(Wp_\tau^*)] = \mathbb{E}\left[u\left(W(1 - \frac{\tau}{T}\lambda - \frac{\tau}{T}\pi)\right)\right]. \tag{10.5}$$

The task now remains to find λ, and hence the final proceeds of bond, for which the agent

[65] For simplicity, no attempt is made to include expected inflation or the Fisher Effect (see Section 6.3) since both assets would be affected equally.

is indifferent between investing in the bond or in cash. This task is equivalent to solving equation (10.5) for $\pi = 0$.

Substituting equation (10.1) into the LHS of equation (10.5) gives[66]

$$\mathbb{E}[u(Wp_\tau^*)] = \mathbb{E}\left[u\left(W\left(1 + B_\tau - \frac{\tau}{T}B_T\right)\right)\right], \tag{10.6}$$

and, by taking a Taylor series expansion about W, equation (10.6) approximates to

$$\begin{aligned} \cong \mathbb{E}\Big[u(W) + W\left(B_\tau - \frac{\tau}{T}B_T\right)u'(W) \\ + \frac{1}{2}W^2\left(B_\tau - \frac{\tau}{T}B_T\right)^2 u''(W)\Big]. \end{aligned} \tag{10.7}$$

Noting that, by the properties of a Brownian bridge,

$$\mathbb{E}\left[B_t - \frac{t}{T}B_T \middle| t = \tau\right] = 0, \tag{10.8}$$

and

$$\mathbb{E}\left[\left(B_t - \frac{t}{T}B_T\right)^2 \middle| t = \tau\right] = \mathbb{E}\left[\sigma^2 t\left(1 - \frac{t}{T}\right) \middle| t = \tau\right], \tag{10.9}$$

and substituting equations (10.8) and (10.9) into (10.7), we arrive at

$$\mathbb{E}[u(Wp_\tau^*)] \cong \mathbb{E}\left[u(W) + \frac{1}{2}W^2\sigma^2\tau\left(1 - \frac{\tau}{T}\right)u''(W)\right]. \tag{10.10}$$

Next, setting $\pi^* \equiv \lambda + \pi$, and assuming that π^* is small, the RHS of equation (10.5) can also be expanded using a Taylor series expansion about W to become

$$\mathbb{E}\left[u\left(W(1 - \frac{\tau}{T}\pi^*)\right)\right] \cong \mathbb{E}\left[u(W) - \frac{\tau}{T}\pi^* W u'(W)\right]. \tag{10.11}$$

Equating the RHS of equation (10.10) with the RHS of equation (10.11) and cancelling the common term $u(W)$ gives

$$\mathbb{E}\left[\frac{1}{2}W^2\sigma^2\tau\left(1 - \frac{\tau}{T}\right)u''(W) + \frac{\tau}{T}\pi^* W u'(W)\right] = 0. \tag{10.12}$$

Taking expectations over the random sale time $\tau \sim U(0, T)$, we have

---

[66] The law of iterated expectations, $\mathbb{E}[x_\tau] = \mathbb{E}\big[\mathbb{E}[x_t|t = \tau]\big]$, will be used implicitly throughout the following derivation.

$$\int_0^T \left\{\frac{1}{2}W^2\sigma^2\left(t-\frac{t^2}{T}\right)u''(W)+\frac{t}{T}\pi^* W u'(W)\right\} f(t)dt = 0. \tag{10.13}$$

Evaluating the integral, with $f(t)=\frac{1}{T}$, yields

$$\left[\left\{\frac{1}{2}W^2\sigma^2\left(\frac{t^2}{2}-\frac{t^3}{3T}\right)u''(W)+\frac{t^2}{2T}\pi^* W u'(W)\right\}\frac{1}{T}\right]_0^T = 0, \tag{10.14}$$

which, over the interval $[0,T]$, becomes

$$\left\{\frac{1}{2}W^2\sigma^2\left(\frac{T^2}{2}-\frac{T^3}{3T}\right)u''(W)+\frac{T^2}{2T}\pi^* W u'(W)\right\}\frac{1}{T} = 0. \tag{10.15}$$

Simplifying, rearranging, recalling that $\pi^* \equiv \lambda + \pi$, and setting $\pi = 0$ gives the general expression for the liquidity premium, or safe rate of interest, as

$$\lambda = -\frac{\sigma^2}{6}\frac{Wu''(W)}{u'(W)}. \tag{10.16}$$

For utility function specified by equation (10.2), the first and second derivatives are

$$u'(W) = W^{-\gamma}, \tag{10.17}$$

and

$$u''(W) = -\gamma W^{-\gamma-1}, \tag{10.18}$$

which, when substituted into equation (10.16), give

$$\lambda = \frac{\gamma\sigma^2}{6}. \tag{10.19}$$

Thus, the liquidity premium required by the agent to be indifferent between money and bonds is proportional to the agent's uncertainty about the rate of return from exchange liquidity σ^2 and the agent's degree of risk aversion γ. In other words, the agent's safe rate of interest is determined by the agent's subjective uncertainty concerning the future rate of interest as postulated by Keynes. This finding is consistent with the interpretation of the theory of liquidity preference expressed in this study and measures the value of the difference between exchange and redemption liquidity. Across a host of agents with a variety of opinions, a spectrum of safe rates of interest will emerge. Any one agent's actual investment allocation will be based on the difference or degree of divergence between the

market price and the price associated with their safe rate of interest (see Section 8.4). At a macroeconomic level, liquidity preference will determine demand-side drivers of the rate of interest.

Similar arguments can be applied to the supply side of the rate of interest. In accordance with the precepts of circuit theory, the relative demand for money and bonds affects the banking system and final finance. The allocation from money to bonds results in a reflux of the opening stock of money in the monetary circuit. The initial flow of money itself depends on the interest rate, both due the elasticity of demand for loans and the asset and liability management policies of the banking system. The balance sheet of the banking system intermediates between borrowers with a tendency to borrow long and investors preferring to lend short. The bank profits from this mismatch, but only by managing the liquidity risk, and, because of the effect that this risk management has in interest rates, the supply and demand for money are not independent.

In borrowing from banks to pay for the factors of production, producers become more illiquid than they would like. They move from their preferred level of liquidity by either spending existing money balances or newly borrowed funds. This deviation from their preferred liquidity position is expected to be temporary because they anticipate the recovery of the proceeds of sale before too long. In a buoyant economy, producers are confident of sales and the liquidity risk is low (see Section 5.4). Similarly, buoyant customers have a reduced demand for deposits, so that the interest rate to satisfy liquidity preference is also low. Banks, if they match their maturity profile, pass their liquidity risk on to borrowers and lenders. In this way, the liquidity provided to the banking system is related to the confidence in near-term sales of producers as part of the theory of the monetary circuit. A fall in confidence in sales, by curtailing the short-term borrowing required to produce saleable goods also removes the natural source of liquidity for the banking system.

Liquidity preference, together with the quasi-rents of assets, determines asset prices. Interest rates are determined by asset allocation decisions between assets with exchange liquidity and assets with redemption liquidity. The propensity to consume out of income reflects time preferences. Therefore, the prices of long-term bonds are not based on expectations as to what will happen over their 30-year term, their prices are based on liquidity discount factors, which are largely derived from uncertainty about discount factors in the near future—the preference of near-term cash flows over the abstract value of a

capital asset. Thirty-year interest rates cannot balance the supply and demand of 30-year investments and saving. Nor can Fisherian time preference explain interest rates. Only liquidity preference of cash flow maturity can explain the pricing of short and long-term assets in the present moment. The expected return on a 30-year asset relates to the value risk inherent in the asset if it needs to be realised for cash in the near future.

As a measure of the degree of preference for redemption liquidity over exchange liquidity, λ is cardinal (and hence ordinal as well). A liquidity premium λ can be derived for each maturity T, and the complex of stochastic discount factors generated enforce the consumption trade-offs set out in the Euler equation (6.6). Instead of the real economy dictating the real rate of interest, as in the neoclassical prescription, the nominal rate of interest is set by liquidity preference and the real economy adjusts to this nominal rate. The Euler equation (6.6) should be matched, not to the marginal productivity of capital, but to the complex of interest rates derived from liquidity preference. An amount saved today can be saved in any form of generic financial asset across a spectrum of maturities, and the rate of interest ensures that the image of value in the unit of account is the same for all. The choice is made on the basis of liquidity preference, where the complex of interest rates is extrapolated from short to long rates. The long rate of interest is a cardinal metric derived from the asset price that ensures indifference between holding an asset for which only market liquidity is available, versus a price-protected asset. This choice of implied interest rate on an asset has no effect on the amount saved.

The image of value in the unit of account arising in exchange is determined by the liquidity discount factor across the spectrum of tenure. The valuation applies regardless of whether the flows are directed into money or long-term illiquid property. The symmetry of exchange is affected by liquidity preference—the ultimate determinant of the term structure of interest rates. Interest rates maintain the symmetry of exchange for generic financial assets and the loosening of the invariance requirement of immediate settlement. This ontological insight reinforces Keynes's observation that interest rates cannot be a reward for saving (see Section 8.1). With the recognition that generic and synthetic financial instruments bring with them entirely new risks, investors have to balance the relative benefits of exchange liquidity and redemption liquidity.

10.8 Summary

The conceptual framework presented here identifies the fundamental distinction between exchange and redemption liquidity, where each is associated with a financial put option that offers liquidity services to its owner. Neither exchange liquidity nor redemption liquidity is an essential property of any specific asset, but are differentiated by the mechanism by which the liquidity option is offered to the asset owner.

Exchange liquidity covers the tradability offered by dealers in an organised market who extract profits by mitigating the costs of search and delay in exchange. These profits are the reward to dealers for providing exchange liquidity on the assets and liabilities of others and translate to a user cost for their customers. Redemption liquidity arises in the specific conditions related to reflux and liability management of money issuers. The theories of the monetary circuit and endogenous money are positioned in the framework as descriptions of the process of liability management and the operation of redemption liquidity.

Once this liquidity dichotomy is identified, the principles underlying the theory of liquidity preference are clarified, and the validity of Keynes's two-asset model can be recognised. Liquidity preference exists on the cusp between exchange and redemption liquidity and operates to balance the relative supply and demand for each. The balancing factor is the rate of interest and it is here that the difference between theory of liquidity preference, based in the tradition of Monetary Analysis, and the interest rate theories of Real Analysis, becomes most apparent. Real Analysis, where perfect exchange liquidity is awarded to all assets, cannot encompass or comprehend this dichotomy. The hypothetical Walrasian auctioneer, by clearing all credit at the beginning of time, entirely abstracts from the risks associated with the change in value of delayed cash flows provided as means of payment or investment. This abstraction from the risks associated with the property of exchange tenure removes any reason to favour redemption liquidity over exchange liquidity.

The risk arising from exchange involving generic financial instruments, and the liquidity dichotomy, can be illustrated using a simple Brownian bridge model. The uncertainty associated with a reliance on exchange liquidity can be compared with the safety of a price-protected asset offering redemption liquidity. In agreement with the theory of liquidity preference, the rate of interest is explained by the compensation required to adopt the risks associated with exchange liquidity.

In the next chapter, this conceptual framework is used to indicate the direction and implications for policy and future research.

11 Conclusion

11.1 Introduction

A clarification of the forms of liquidity and the nature of money in Part II formed the basis of an immanent critique of the literature on liquidity and money in Part III and Part IV of this study. In Part III, it was shown that the treatment of liquidity and money within the tradition of Real Analysis fails to achieve the level of correspondence with reality required to be consistent with the realist methodology adopted by this study. Consequently, the insights gained from this literature are of limited relevance for policy.

In Part IV of the study, the critique led to an extension and new interpretation of Keynes's theory of liquidity preference. The interpretation builds on, but is significantly different from, the previous interpretations of liquidity preference in the literature, such as those of Bronfenbrenner, Hicks and Tobin, by focusing on liquidity and not market risk.

Two related contributions are made to knowledge by the conceptual framework constructed in Part V. First, the distinction between redemption and exchange liquidity reduces the many forms of liquidity into two distinct categories and thereby clarifies the relationship between liquidity and money. Second, the distinction between redemption and exchange liquidity forms a basis for a new interpretation of Keynes's theory of liquidity preference as a determinant of the rate of interest.

11.2 Summary: Addressing the Research Questions

Money is the result of a synthetic financial exchange, and as such exists as a social relation represented by balance-sheet transactions. Being created as act of synthetic exchange, money has no endogenous limit and its value is based on an underlying reference object, which varies by institutional and historical context. Such examples are: central bank money based on a gold standard; private sector banks that maintain convertibility into a central bank or state money; or fiat money with reference to a consumer price index. None of these reference objects applies any limit to the creation of synthetic money, although the anchoring to the reference object does create the need to manage the resulting stock of money by the adjustment of interest rates.

With regard to the first research question posed in Section 1.2, the relationship between liquidity and money is dominated by the ontological view that one takes of money. When

money is viewed as the most tradable commodity as discussed in Section 3.3, then liquidity is reduced to saleability, and, in the neo-Walrasian benchmark model, money is eliminated altogether. As shown in Part III, Real Analysis, whereby money is treated as a commodity, fails to incorporate liquidity and instead focuses on market liquidity and transactions costs. Neither the neo-Walrasian nor the Wicksellian forms of general equilibrium theory are therefore suitable to generate insights into the nature of liquidity. Wicksellian models, due to their reliance on the marginal productivity of capital, are theoretically inconsistent. Neo-Walrasian pure-exchange models, lacking money, fail to achieve sufficient correspondence with reality to build a coherent framework to consider liquidity.

By contrast, if money is viewed as a social relationship based on balance-sheet debits and credits (see Section 3.6) and it is recognised that the tenure of exchange can extend beyond the immediate settlement of commodity-for-commodity contracts (see Section 8.6), then the concept of liquidity can be placed into a temporal dimension associated with preferences for credit clearing. Specifically, this temporal dimension allows for the distinction between exchange liquidity and redemption liquidity. Exchange liquidity represents the transfer of financial assets between counterparties unrelated to the issuer of the financial asset, whereas redemption liquidity refers to the repayment option embedded in money whereby it can be used to cancel debts with the money's issuer by the mechanism of reflux.

To answer the second research question, the distinction between redemption and exchange liquidity forms the basis of the conceptual framework of liquidity presented in Section 10.3. Importantly, liquidity is recognised as not being an intrinsic property of an asset but, instead, is a form of financial contract provided by liquidity suppliers. Issuers of money must manage the reflux mechanism by adjusting the rate of interest paid on their non-redeemable liabilities. By contrast, the demand for liquidity arises as the result of the trade-off between the certainty of redemption liquidity and the uncertainty of relying on the price risk associated with exchange liquidity. This trade-off between redemption and exchange liquidity forms the basis of a new interpretation of Keynes's theory of liquidity preference where the current rate of interest is determined by uncertainty regarding the future rate of interest. The rate of interest, then, is not the balancing factor between real investment and real saving as it is in the theory of loanable funds. Instead, it equilibrates the competing preferences for credit assets arising through the extended tenure of exchange. This insight

suggests that the construction of theory with regard to the rate of interest must be conducted from within the tradition of Monetary Analysis, not Real Analysis.

11.3 Limitations

Due to the scope of this study—the literature related to liquidity and money is vast—the analysis is broad rather than deep. The focus has pragmatically been on the area of the literature that reveals the ontological assumptions underlying each theory of liquidity and money. The framework organises these ontological concepts but stops short of any theoretical or empirical analysis.

A further limitation is that, although the existence of uncertainty is essential to the explanation of liquidity and money, the concept of uncertainty itself has conflicting definitions in the literature. Clarifications of uncertainty are beyond the scope of this study, and therefore questions concerning its ontology and epistemology are left open. Further research is required to clarify its relationship with, for example, non-ergodicity and chaos theory.

11.4 Implications for Policy

The explanation of the rate of interest by factors related to exchange and redemption liquidity suggests reasons for the failure of the global economy to revert to growth paths experienced before the Global Financial Crisis (GFC). The preoccupation of policy makers and academics with the Wicksellian natural rate of interest has informed such explanations as secular stagnation or a savings glut and is behind the concerns regarding the 'artificial' stimulus provided by unorthodox central-bank monetary policy. That massive monetary creation has not generated inflation is also a mystery for those whose conceptual framework maintains a place for the quantity theory of money. The behaviour of the global economy over the past decade has provided counter-examples for both the quantity theory of money and the inevitable alignment of central bank rates with a rate of interest determined purely in the real economy. Interest rate theories anchored in Real Analysis fail to explain the low level of interest rates prevalent beyond a period of time that can sensibly be aligned with anything other than the long run.

As shown in Part IV, the Keynesian and post-Keynesian theories of liquidity preference, endogenous money and the monetary circuit provide explanations consistent with observed

financial market behaviour, both before and after the GFC. Low interest rates are now conventional in the sense that the economy has adapted to them, and consequently the determination expressed by the monetary authorities to 'normalise' policy and balance-sheet size would now be highly disruptive if enacted. That low rates have not provided the launch pad for an investment boom indicates that Keynes's two-price mechanism driving investment is not a latent factor suppressed by uncertainty. Instead, uncertainty regarding the future rate of interest determines the current rate of interest, and policies such as QE satisfy a liquidity preference, but do not create spending. The circular flow of income and a desire to hold liquid stores of value are separate phenomena. Central-bank policy makers are trapped in the confusion generated by their inappropriate analytical framework.

The trade-off between redemption and exchange liquidity is most apparent in repo, which formed the basis of the shadow banking system. The security of collateral and the application of a haircut to its value to protect the money lender mean that repo is almost entirely free from credit and market risk, leaving the exchange as a pure expression of liquidity preference. The framework offered by this study is suitable for a deeper consideration of the shadow banking system, its potential benefits and the reasons for its failure.

The distinction between exchange and redemption liquidity, coupled with an emphasis on collateral, provides a better understanding of yield-curve dynamics. For example, shortages in redemption liquidity, which raise short-term funding rates, can also increase the demand (that is, lower the yield) for long-term assets that serve as collateral for repo-based short-term funding (Pozsar, 2017). Thus, it is possible to understand the slope of the yield curve entirely in terms of current redemption liquidity provision, and not expectations of future short-term rates. Monetary authorities will be able use this framework to interpret market movements.

Several policy implications of this framework directly concern the banking system. One area of consideration, discussed in Section 7.3, is the push to eliminate 'run-prone securities' from the banking system. The successful application of this initiative would be to limit the financial system to issuing assets subject to exchange liquidity and eliminate the provision of redemption liquidity. Such a policy approach is based on ontological confusion as to the nature of bank liabilities coupled with a misreading of the purpose and innovation underling the shadow banking system. In its attempt to provide redemption

liquidity, the phenomenon of shadow banking shows two things. First, the significance of the trade-off between redemption and exchange liquidity means that financial innovation partly consists of an unceasing effort to convert exchange liquidity into redemption liquidity. Any attempt to restrict redemption liquidity in one part of the system will result in an attempt to revive it elsewhere. Second, and more important, the shadow banking system demonstrated that the provision of redemption liquidity is unstable and requires the support of a lender or dealer of last resort. It would be unwise to push its provision into parts of the financial system without access to this assistance.

11.5 Implications for Future Research

Many empirical and theoretical avenues for future research are open. First, the importance of collateral has been amplified in the post-GFC environment, with the consequence that the supply and valuation of collateral have the potential to affect significantly the provision and cost of liquidity. Second, more research is needed to investigate the macroeconomic significance of the safe rate of interest (see Sections 8.1 and 10.7), as well as empirical research into the link between uncertainty and the entire complex of interest rates. Third, the distinction between exchange and redemption liquidity has implications for the interpretations of the risk-free rate and the market-risk premium in the Capital Asset Pricing Model (CAPM). A deeper understanding of exchange liquidity could also be used to assess the benefits of exchange-traded funds (ETFs). Finally, the framework could be used to answer the question of whether the current environment of low interest rates globally is in accord with Keynes's policy of cheap money and perhaps heralds the 'euthanasia of the rentier'.

The explicit linkage between redemption and exchange liquidity, uncertainty and the rate of interest offers macroeconomic modellers a better alternative to the incoherent theories founded on time preference and the marginal productivity of capital (Rogers, 1989). Post-Keynesian stock-flow consistent (SFC) models (Godley & Lavoie, 2012) often rely on portfolio-rebalancing models and asset stocks (see Section 8.3), and hence suffer from asset fetishism (see Section 8.6). SFC models could be improved by the pricing of risk and liquidity outlined by the conceptual framework offered here. Linking the rate of interest (or the demand for redemption liquidity) to uncertainty would allow the models to reflect a greater level of realism in their treatment of money creation, the operation of the banking

system and asset returns.

11.6 Concluding Note

This work goes a long way towards clarifying the confusion surrounding the concept of liquidity. A taxonomy of liquidity is offered, and a crucial distinction between redemption and exchange liquidity is identified. This distinction allows for a new interpretation of Keynes's theory of liquidity preference that provides insights into the current economic environment, an understanding of recent innovations in finance, and explanations for the failure of the shadow banking system. The importance of the tradition of Monetary Analysis, as the appropriate method of explaining and understanding liquidity, is established. This research also provides a basis for many different strands of future research in areas as diverse as: asset pricing; monetary policy and financial stability; the effect of collateral on the post-GFC financial system; the future of shadow banking; and the importance of public investment, in comparison with monetary policy, as a means of economic stimulus.

References

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics, 77*(2), 375-410.

Admati, A. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica, 53*(3), 629-658.

Admati, A., & Hellwig, M. (2013). *The bankers' new clothes: What's wrong with the banking system and what to do about it*. Princeton, NJ: Princeton University Press.

Adrian, T., & Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation, 19*(3), 418-437.

Ahamed, L., & Bernanke, B. (2014). A conversation with Ben Bernanke. In D. Wessel (Ed.), *Central banking after the great recession: Lessons learned and the challenges ahead* (pp. 5-20). Washington DC: Brookings Institution Press.

Albanese, C., Caenazzo, S., & Crepey, S. (2016, March 12). *Capital valuation adjustment and funding valuation adjustment*. Retrieved from https://ssrn.com/abstract=2745909

Allen, F., & Gale, D. (1998). Optimal financial crises. *The Journal of Finance, 53*(4), 1245-1284.

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of Political Economy, 108*(1), 1-33.

Allen, F., & Gale, D. (2004). Financial fragility, liquidity, and asset prices. *Journal of the European Economic Association, 2*(6), 1015-1048.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets, 5*(1), 31–56.

Amihud, Y., Mendelson, H., & Pedersen, L. H. (2005). Liquidity and asset prices. *Foundations and Trends in Finance, 1*(4), 269-364.

Arestis, P., & Howells, P. (1996). Theoretical reflections on endogenous money: the problem with 'convenience lending'. *Cambridge Journal of Economics, 20*(5), 539-551.

Arrow, K., & Hahn, F. (1971). *General competitive analysis*. San Francisco, CA: Holden Day.

Bagehot, W. (1873). *Lombard street: A description of the money market*. Retrieved from http://www.econlib.org

Barro, R. J. (1974). Are government bonds net wealth? *Journal of Political Economy*, *82*(6), 1095-1117.

Beggs, M. (2012, March). *Liquidity as a social relation*. Paper presented at the Annual Meeting of the Eastern Economic Association, Boston, MA.

Beggs, M. (2015, June). *Liquidity's other career*. Paper presented at the Annual Meeting of the History of Economics Society, Lansing, MI.

Beggs, M. (2016). The state as a creature of money. *New Political Economy, 22*(5), 463-477.

Bell, S. (2000). Do taxes and bonds finance government spending? *Journal of Economic Issues*, *34*(3), 603-620.

Bell, S. (2001). The role of the state and the hierarchy of money. *Cambridge Journal of Economics*, *25*(2), 149-163.

Bernanke, B. (2005, March 10). *The global saving glut and the US current account deficit*. Speech at the Sandridge Lecture, Virginia Association of Economics, Richmond, Virginia.

Bibow, J. (1995). Some reflections on Keynes's 'finance motive' for the demand for money. *Cambridge Journal of Economics*, *19*(5), 647-666.

Bibow, J. (1998). On Keynesian theories of liquidity preference. *The Manchester School*, *66*(2), 238-273.

Bibow, J. (2000). On exogenous money and bank behaviour: The Pandora's box kept shut in Keynes' theory of liquidity preference? *The European Journal of the History of Economic Thought*, *7*(4), 532-568.

Bibow, J. (2001). The loanable funds fallacy: Exercises in the analysis of disequilibrium. *Cambridge Journal of Economics*, *25*(5), 591-616.

Bibow, J. (2005). Liquidity preference theory revisited—to ditch or to build on it? *Levy Economics Institute Working Papers Series No. 427.*

Bibow, J. (2009). *Keynes on monetary policy, finance and uncertainty: Liquidity*

preference theory and the global financial crisis. London, UK: Routledge.

Black, F. (1970a). *Fundamentals of liquidity*. Graduate School of Business, University of Chicago. Retrieved from http://www.rasmusen.org/special/black/FundamentalsofLiquidity70.pdf

Black, F. (1970b). Banking and interest rates in a world without money. *Journal of Bank Research, 1*(9), 8-20.

Black, F. (1971). Toward a fully automated stock exchange, part I. *Financial Analysts Journal, 27*(4), 28-35.

Black, F. (1986). Noise. *Journal of Finance, 41*(3), 529-543.

Black, F. (1987). *Business cycles and equilibrium*. Oxford, UK: Basil Blackwell.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy, 81*(3), 637-654.

Block, W. E., & Barnett II, W. (2012). Milton Friedman and the financial crisis of 2008. *American Review of Political Economy, 10*(1/2), 2.

Boulding, K. E. (1944). A liquidity preference theory of market prices. *Economica, 11*(42), 55–63.

Bolton, P., Santos, T., & Scheinkman, J. A. (2011). Outside and inside liquidity. *The Quarterly Journal of Economics, 126*(1), 259-321.

Botta, A., Caverzasi, E., & Tori, D. (2015). Financial–real-side interactions in an extended monetary circuit with shadow banking: Loving or dangerous hugs? *International Journal of Political Economy, 44*(3), 196-227.

Boumans, M., & Davis, J. (2010). *Economic methodology: Understanding economics as a science*. Basingstoke, UK: Palgrave Macmillan.

Bronfenbrenner, M. (1945). Some fundamentals in liquidity theory. *Quarterly Journal of Economics, 59*(3), 405-26.

Brown, C. (2003). Toward a reconcilement of endogenous money and liquidity preference. *Journal of Post Keynesian Economics, 26*(2), 325-340.

Brown, S. J. (2011). The efficient markets hypothesis: The demise of the demon of chance? *Accounting and Finance, 51*(1), 79-95.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *The Journal of Economic Perspectives*, *23*(1), 77-100.

Brunnermeier, M. K., Eisenbach, T. M., & Sannikov, Y. (2013). Macroeconomics with financial frictions: A survey. In D. Acemoglu, M. Arellano & E. Dekel (Eds.), *Advances in economics and econometrics, tenth world congress of the Econometric Society* (Vol. 2, pp. 3-94). New York, NY: Cambridge University Press.

Brunnermeier, M. K., Gorton, G., & Krishnamurthy, A. (2011). Risk topography. *NBER Macroeconomics Annual*, *26*(1), 149-176.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, *22*(6), 2201-2238.

Buiter, W. H. (2005). New developments in monetary economics: Two ghosts, two eccentricities, a fallacy, a mirage and a mythos. *The Economic Journal*, *115*(502), C1-C31.

Buiter, W. H. (2009, March 6). The unfortunate uselessness of most 'state of the art' academic monetary economics. Retrieved from http://ssrn.com/abstract=2492949

Buiter, W. H. (2014, August 21). *The simple analytics of helicopter money: Why it works – always*. http://dx.doi.org/10.5018/economics-ejournal.ja.2014-28

Cannan, E. (1921). The meaning of bank deposits. *Economica*, *1*(1), 28-36.

Cardim de Carvalho, F. J. (2010). Uncertainty and money: Keynes, Tobin and Kahn and the disappearance of precautionary demand for money from liquidity preference. *Cambridge Journal of Economics*, *34*(4), 709-725.

Cass, D., & Shell, K. (1983). Do sunspots matter? *Journal of Political Economy*, *91*(2), 193-227.

Castagna, A. (2012, August 30). *Yes, FVA is a cost for derivatives desks - a note on 'Is FVA a cost for derivatives desks?' by Prof. Hull and Prof. White*. Retrieved from https://ssrn.com/abstract=2141663

Chick, V. (1983). *Macroeconomics after Keynes: A reconsideration of the General Theory*. Cambridge, MA: The MIT Press.

Chick, V. (1991). Hicks and Keynes on liquidity preference: A methodological approach.

Review of Political Economy, *3*(3), 309-319.

Chick, V. (1992). The evolution of the banking system and the theory of saving, investment and interest. In P. Arestis & S. C. Dow (Eds.), *On money, method and Keynes: Selected essays: Victoria Chick* (pp. 193-205). Basingstoke, UK: Macmillan Press.

Chick, V., & Dow, S. (2002). Monetary policy with endogenous money and liquidity preference: A nondualistic treatment. *Journal of Post Keynesian Economics*, *24*(4), 587-607.

Chick, V., & Tily, G. (2014). Whatever happened to Keynes's monetary theory? *Cambridge Journal of Economics*, *38*(3), 681-699.

Chordia, T., Sarkar, A., & Subrahmanyam, A. (2005). An empirical analysis of stock and bond market liquidity. *Review of Financial Studies*, *18*(1), 85–129.

Clower, R. (1967). A reconsideration of the microfoundations of monetary theory. *Western Economic Journal*, *6*(1), 1-8.

Cochrane, J. H. (1992). Explaining the variance of price-dividend ratios. *The Review of Financial Studies*, *5*(2), 243-280.

Cochrane, J. H. (2005a). Asset pricing program review: Liquidity, trading and asset prices. *NBER Reporter*, *Winter 2004/5*.

Cochrane, J. H. (2005b). *Asset pricing.* Princeton, NJ: Princeton University Press.

Cochrane, J. H. (2005c). Money as stock. *Journal of Monetary Economics*, *52*(3), 501-528.

Cochrane, J. H. (2011a). Determinacy and identification with Taylor rules. *Journal of Political Economy*, *119*(3), 565-615.

Cochrane, J. H. (2011b). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047-1108.

Cochrane, J. H. (2014). Toward a run-free financial system. In M. N. Baily & J. B. Taylor (Eds.), *Across the great divide: New perspectives on the financial crisis* (pp. 197-249). Stanford, CA: Hoover Institution Press.

Cochrane, J. H. (2017a). Macro-finance. *Review of Finance*, *21*(3), 945-985

Cochrane, J. H. (2017b). Michelson-Morley, Fisher, and Occam: The radical implications

of stable quiet inflation at the zero bound. In *NBER Macroeconomics Annual 2017*, *32*.

Conard, J. W. (1959). *Introduction to the theory of interest*. Los Angeles, CA: University of California Press.

Copeland, T. E., & Galai, D. (1983). Information effects on the bid–ask spread. *Journal of Finance*, *38*(5), 1457–1469.

Cottrell, A. (1994a). Post-Keynesian monetary economics. *Cambridge Journal of Economics*, *18*(6), 587–605.

Cottrell, A. (1994b). Keynes and the Keynesians on the Fisher effect. *Scottish Journal of Political Economy*, *41*(4), 416-433.

Cottrell, A. (1997). Monetary endogeneity and the quantity theory: The case of commodity money. *HOPE*.

Crotty, J. (2011). The realism of assumptions does matter: Why Keynes-Minsky theory must replace efficient market theory as the guide to financial regulation policy. In M. H. Wolfson & G. A. Epstein (Eds.), *The handbook of the political economy of financial crises* (pp. 133-158). Oxford, UK: Oxford University Press.

Dang, T. V., Gorton, G., & Holmström, B. (2012). Ignorance, debt and financial crises. *Yale University and Massachusetts Institute of Technology Working Paper No. 17*.

Davidson, P. (1965). Keynes's finance motive. *Oxford Economic Papers, 17*(1), 47-65.

Davidson, P. (1987). Sensible expectations and the long-run non-neutrality of money. *Journal of Post Keynesian Economics, 10*(1), 146-153.

Davidson, P. (1988). A technical definition of uncertainty and the long-run non-neutrality of money. *Cambridge Journal of Economics, 12*(3), 329-337.

Davidson, P. (2006). Keynes and money. In P. Arestis & M. Sawyer (Eds.), *A handbook of alternative monetary economics* (pp. 139-153). Cheltenham, UK: Edward Elgar.

Davidson, P. (2008). Securitization, liquidity, and market failure. *Challenge*, *51*(3), 43-56.

Davidson, P. (2015). A rejoinder to O'Donnell's critique of the ergodic/nonergodic explanation of Keynes's concept of uncertainty. *Journal of Post Keynesian Economics, 38*(1), 1-18.

Davidson, P., & Weintraub, S. (1973). Money as cause and effect. *The Economic Journal, 83*(332), 1117-1132.

Davies, G. (2002). *A history of money: From ancient times to the present day*. Cardiff, UK: University of Wales Press.

Davies, H. (2010). *The financial crisis: Who is to blame?* Cambridge, UK: Polity.

Davis, A. (1900). Currency and banking in the province of the Massachusetts-Bay: Part I: Currency. *Publications of the American Economic Association, 3rd Series, 1*(4), 1-473.

De Cecco, M. (1991). Gold standard. In J. Eatwell, M. Milgate & P. Newman (Eds.), *The new Palgrave: The world of economics* (pp. 315-329). New York, NY: Macmillan Press.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*. New Haven, CT: Yale University Press.

Dent, A., & Dison, W. (2012). The Bank of England's Real-Time Gross Settlement infrastructure. *Bank of England Quarterly Bulletin*, Q3.

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *The Journal of Political Economy, 91*(3), 401-419.

Dow, S. C. (1996). Horizontalism: A critique. *Cambridge Journal of Economics, 20*(4), 497-508.

Dunbar, C. F. (1922). *The theory and history of banking* (4th ed.). New York, NY: G. P. Putnam's Sons.

Einzig, P. (1954). *How money is managed: The ends and means of monetary policy*. Middlesex, UK: Penguin Books Ltd.

Elsner, W., Heinrich, T., & Schwardt, H. (2014). *The microeconomics of complex economies: Evolutionary, institutional, neoclassical, and complexity perspectives*. Oxford, UK: Elsevier.

Espinoza, R. A., Goodhart, C. A. E., & Tsomocos, D. P. (2009). State prices, liquidity, and default. *Economic Theory, 39*(2), 177-194.

Ertürk, K. (2006). Speculation, liquidity preference, and monetary circulation. In P. Arestis & M. Sawyer (Eds.), *A handbook of alternative monetary economics* (pp. 454-470). Cheltenham, UK: Edward Elgar.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417.

Fama, E. F. (1980). Banking in the theory of finance. *Journal of Monetary Economics*, *6*(1), 39-57.

Fama, E. F. (1991). Efficient capital markets II. *The Journal of Finance*, *46*(5), 1575-1617.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3-25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23-49.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*(1), 55-84.

Ferguson, N. (2008). *The ascent of money: A financial history of the world*. Australia: Penguin.

Fisher, I. (1896). Appreciation and interest. *Publications of the American Economic Association*, *11*(4), 1-98.

Foley, D. (2005). Marx's theory of money in historical perspective. In F. Moseley (Ed.), *Marx's theory of money: Modern appraisals* (pp. 65-77). Basingstoke, UK: Palgrave Macmillan.

Fontana, G. (2000). Post Keynesians and circuitists on money and uncertainty: An attempt at generality. *Journal of Post Keynesian Economics*, *23*(1), 27-48.

Fontana, G. (2004). Rethinking endogenous money: A constructive interpretation of the debate between horizontalists and structuralists. *Metroeconomica*, *55*(4), 367–385.

Friedman, M. (1953a). The methodology of positive economics. In M. Friedman (Ed.), *Essays in positive economics* (pp. 3-43). Chicago, IL: Chicago University Press.

Friedman, M. (1953b). The case for flexible exchange rates. In M. Friedman (Ed.), *Essays in positive economics* (pp. 157-203). Chicago, IL: Chicago University Press.

Friedman. M. (1956). The quantity theory of money—A re-statement. In M. Friedman (Ed.), *Studies in the Quantity Theory of Money* (pp. 3-21). Chicago, IL: Chicago

University Press.

Fullwiler, S. T. (2003). Timeliness and the Fed's daily tactics. *Journal of Economic Issues*, *37*(4), 851-880.

Gabor, D., & Vestergaard, J. (2016). Towards a theory of shadow money. *Institute of New Economic Thinking (INET) Working Paper*.

Galbraith, J. K. (1975). *Money: Whence it came, where it went*. London, UK: Penguin.

Galí, J. (2008). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework*. Princeton, NJ: Princeton University Press.

Geanakoplos, J. (1990). An introduction to general equilibrium with incomplete asset markets. *Journal of Mathematical Economics*, *19*(1), 1-38.

Gertler, M., & Kiyotaki, N. (2011). Financial intermediation and credit policy in business cycle analysis. In B. M. Friedman & M. Woodford (Eds.), *Handbook of monetary economics* (Vol. 3A, pp. 547-599). San Diego, CA: Elsevier B. V.

Glasner, D. (1985). A reinterpretation of classical monetary theory. *Southern Economic Journal*, *52*(1), 46-67.

Glasner, D. (1989). On some classical monetary controversies. *History of Political Economy*, *21*(2), 201-229.

Glasner, D. (2000). Classical monetary theory and the quantity theory. *History of Political Economy*, *32*(1), 39-59.

Godley, W., & Lavoie, M. (2012). *Monetary economics: An integrated approach to credit, money, income, production and wealth* (2nd ed.). Basingstoke, UK: Palgrave Macmillan

Goodhart, C. A. E. (1989). Has Moore become too horizontal? *Journal of Post Keynesian Economics*, *12*(1), 29-34.

Goodhart, C. A. E. (1991). Is the concept of an equilibrium demand for money meaningful? A reply to "has the demand for money been mislaid?" *Journal of Post Keynesian Economics*, *14*(1), 134-136.

Goodhart, C. A. E. (2005a). Credit and state theories of money: The contributions of A. Mitchell Innes. *History of Political Economy*, *37*(4), 759-761.

Goodhart, C. A. E. (2005b). What can academics contribute to the study of financial stability? *Economic and Social Review*, *36*(3), 189-203.

Goodhart, C. (2008). Liquidity risk management. *Banque de France, Financial Stability Review*, *11*(1), 39-44.

Goodhart, C. A. E. (2009). The continuing muddles of monetary theory: A steadfast refusal to face facts. *Economica*, *76*(s1), 821-830.

Gorton, G., & Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial Economics*, *104*(3), 425-451.

Graeber, D. (2012). *Debt: The first 5,000 years*. Brooklyn, NY: Melville House.

Graziani, A. (1990). The theory of the monetary circuit. *Économies et Sociétés: Monnaie et Production*, *24*(7), 7–36.

Grossman, S. J. (1981). An introduction to the theory of rational expectations under asymmetric information. *The Review of Economic Studies*, *48*(4), 541-559.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*(3), 393–408.

Gurley, J. G., & Shaw, E. S. (1960). *Money in a theory of finance*. Washington, DC: Brookings Institution.

Hahn, F. H. (1982). *Money and inflation*. Oxford, UK: Blackwell.

Hahn, F. H. (1990). Liquidity. In B. M. Friedman & F. H. Hahn (Eds.), *Handbook of Monetary Economics* (Vol. 1, pp. 63-80). Amsterdam, The Netherlands: Elsevier Science B. V.

Hansen, A. H. (1953). *A guide to Keynes*. New York, NY: McGraw-Hill.

Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, *20*(3), 381-408.

Hawtrey, R. G. (1919). *Currency and credit*. London, UK: Longmans, Green and Co.

Hayek, F. A. von (1937). Economics and knowledge. *Economica*, *4*(13), 33-54.

Hayes, M. G. (2006). Value and probability. *Journal of Post Keynesian Economics*, *28*(3), 527-538.

Hayes, M. G. (2010). The loanable funds fallacy: Saving, finance and equilibrium. *Cambridge Journal of Economics, 34*(4), 807-820.

Hirshleifer, J. (1967). A note on the Bohm-Bawerk/Wicksell theory of interest. *The Review of Economic Studies, 34*(2), 191-199.

Hewitson, G. (1993). An intellectual history of money endogeneity theory. *History of Economics Review, 20*(1), 140-160.

Hicks, J. R. (1937). Mr. Keynes and the 'classics': A suggested interpretation. *Econometrica, 5*(2), 147-159.

Hicks, J. R. (1962). Liquidity. *The Economic Journal, 72*(288), 787-802.

Hicks, J. (1974). *The crisis in Keynesian economics*. Oxford, UK: Basil Blackwell.

Hicks, J. (1989). *A market theory of money*. Oxford, UK: Clarendon Press.

Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics, 9*(1), 47-73.

Holmström, B., & Tirole, J. (1998). Private and public supply of liquidity. *Journal of Political Economy, 106*(1), 1-40.

Holmström, B., & Tirole, J. (2001). LAPM: A liquidity-based asset pricing model. *The Journal of Finance, 56*(5), 1837-1867.

Howells, P. G. A. (1995). The demand for endogenous money. *Journal of Post Keynesian Economics, 18*(1), 89-106.

Howells, P. G. A. (1997). The demand for endogenous money: A rejoinder. *Journal of Post Keynesian Economics, 19*(3), 429-435.

Hull, J., & White, A. (2012). Is FVA a cost for derivatives desks? *Risk, 25*(9), 83-85.

Hume, D. (1742). *Essays, moral, political, and literary*. Retrieved from http://www.econlib.org

Ingham, G. (2000). The credit theory of money: The monetary circuit approach. In J. Smithin (Ed.), *What is money?* (pp. 16-41). London, UK: Routledge.

Ingham, G. (2004a). *The nature of money*. Cambridge, UK: Polity.

Ingham, G. (2004b). The emergence of capitalist credit money. In Wray, L. R. (Ed.),

Credit and state theories of money: The contributions of A. Mitchell Innes (pp. 173-222). Cheltenham, UK: Edward Elgar.

Innes, A. M. (1913). What is money? *Banking Law Journal*, *30*(5), 377-408.

Innes, A. M. (1914). The credit theory of money. *Banking Law Journal*, *31*(2), 151-168.

Jakab, Z., & Kumhof, M. (2015). Banks are not intermediaries of loanable funds - and why this matters. *Bank of England Working Paper No. 529*.

Jefferis, C. (2017). *The dialectics of liquidity crisis: An interpretation of the explanations of the financial crisis of 2007-08*. Abingdon, UK: Routledge.

Jespersen, J. (2009). *Macroeconomic methodology: A post-Keynesian perspective*. Cheltenham, UK: Edward Elgar.

Jevons, W. S. (1875). *Money and the mechanism of exchange*. London, UK: Appleton.

Kahn, R. F. (1954). Some notes on liquidity preference. *The Manchester School*, *22*(3), 229-257.

Kaldor, N. (1939). Speculation and economic stability. *The Review of Economic Studies*, *7*(1), 1-27.

Kaldor, N. (1982). *The scourge of monetarism*. Oxford, UK: Oxford University Press.

Kaldor, N., & Trevithick, J. (1981). A Keynesian perspective on money. *Lloyds Bank Review*, *139*(1), 1-19.

Karatzas, I., & Shreve, S. E. (1988). *Brownian motion and stochastic calculus* (2nd ed.). New York, NY: Springer-Verlag.

Kay, J. (2015, November 17). For the ordinary saver, liquidity is of little value. *Financial Times*. Retrieved from http://on.ft.com/1N9wN0M

Keynes, J. M. (1914). Reviewed work: What is money? By A. Mitchell Innes. *The Economic Journal*, *24*(95), 419-21.

Keynes, J. M. (1921). *A treatise on probability*. London, UK: Macmillan and Co.

Keynes, J. M. (1930a). *A treatise on money: Volume I: The pure theory of money*. London, UK: Macmillan.

Keynes, J. M. (1930b). *A treatise on money: Volume II: The applied theory of money*.

London, UK: Macmillan.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London, UK: Macmillan.

Keynes, J. M. (1937a). The general theory of employment. *The Quarterly Journal of Economics*, *51*(2), 209-223.

Keynes, J. M. (1937b). Alternative theories of the rate of interest. *The Economic Journal*, *47*(186), 241-252.

Keynes, J. M. (1937c). The 'ex ante' theory of the rate of interest. *The Economic Journal*, *47*(188), 663-669.

Keynes, J. M. (1937d). The theory of the rate of interest. In A. D. Gayer (Ed.), *The lessons of monetary experience: Essays in honor of Irving Fisher* (pp. 145-152). London, UK: George Allen & Unwin Ltd.

Keynes, J. M. (1979). *The collected writings of John Maynard Keynes, volume XXIX: The general theory and after: A supplement*. London, UK: Macmillan.

King, J. E. (2012). *The microfoundations delusion: Metaphor and dogma in the history of macroeconomics*. Cheltenham, UK: Edward Elgar.

King, M. (1999). Challenges for monetary policy: New and old. *Bank of England Quarterly Bulletin*, *39*(4), 397-415.

King, M. (2016). *The end of alchemy: Money, banking and the future of the global economy*. London, UK: Little, Brown.

Kiyotaki, N., & Wright, R. (1989). On money as a medium of exchange. *The Journal of Political Economy*, *97*(4), 927-954.

Kiyotaki, N., & Wright, R. (1993). A search-theoretic approach to monetary economics. *The American Economic Review*, *83*(1), 63-77.

Klebaner, F. C. (1998). *Introduction to stochastic calculus with applications*. London, UK: Imperial College Press.

Knapp, G. F. (1924). *The state theory of money*. London, UK: Macmillan & Company.

Kregel, J. (1988). The multiplier and liquidity preference: Two sides of the theory of effective demand. In A. Barrère (Ed.), *The foundations of Keynesian analysis* (pp. 231-

250). Basingstoke, UK: The Macmillan Press.

Kregel, J. (2014). Liquidity preference and the entry and exit to ZIRP and QE. *Levy Economics Institute Policy Note No. 2014/15*.

Krugman, P. R. (1998). It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity, 1998*(2), 137-187.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica, 53*(6), 1315-1335.

Laubach, T., & Williams, J. C. (2003). Measuring the natural rate of interest. *Review of Economics and Statistics, 85*(4), 1063-1070.

Lavoie, M. (1984). The endogenous flow of credit and the post Keynesian theory of money. *Journal of Economic Issues, 18*(3), 771-797.

Lavoie, M. (1999). The credit-led supply of deposits and the demand for money: Kaldor's reflux mechanism as previously endorsed by Joan Robinson. *Cambridge Journal of Economics, 23*(1), 103-113.

Lavoie, M. (2006). Endogenous money: Accommodationist. In P. Arestis & M. Sawyer (Eds.), *A handbook of alternative monetary economics* (pp. 17-34). Cheltenham, UK: Edward Elgar.

Lavoie, M. (2010). Changes in central bank procedures during the subprime crisis and their repercussions on monetary theory. *International Journal of Political Economy, 39*(3), 3-23.

Lavoie, M. (2013). The monetary and fiscal nexus of neo-chartalism: A friendly critique. *Journal of Economic Issues, 47*(1), 1-32.

Lavoie, M. (2014). *Post-Keynesian economics: New foundations*. Cheltenham, UK: Edward Elgar.

Lawson, T. (1997). *Economics and reality*. London, UK: Routledge.

Lawson, T. (2015). *Essays on the nature and state of modern economics*. Abingdon, UK: Routledge.

Le Bourva, J. (1992). Money creation and credit multipliers. *Review of Political Economy, 4*(4), 447-466.

Leijonhufvud, A. (1981). *Information and coordination: Essays in macroeconomic theory*. Oxford, UK: Oxford University Press.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics, 47*(1), 13-37.

Lippman, S. A., & McCall, J. J. (1986). An operational measure of liquidity. *The American Economic Review, 76*(1), 43-55.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics, 82*(3), 631-671.

Ljungqvist, L., & Sargent, T. J. (2004). *Recursive macroeconomic theory* (2nd ed.). Cambridge, MA: The MIT Press.

Lowenstein, R. (2001). *When genius failed: The rise and fall of Long-Term Capital Management*. London, UK: Fourth Estate.

Lozano, B. (2015). *Of synthetic finance: Three essays of speculative materialism*. Abingdon, UK: Routledge.

Lucas, R. E., Jr. (1978). Asset prices in an exchange economy. *Econometrica, 46*(6), 1429-1445.

Lucas, R. E., Jr. (1984). Money in a theory of finance. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 21, pp. 9-46).

Lucas, R. E., Jr. (1990). Liquidity and interest rates, *Journal of Economic Theory, 50*(2), 237-264.

Lucas, R. E., Jr. (1996). Nobel lecture: Monetary neutrality. *Journal of Political Economy, 104*(4), 661-682.

Lucas, R. E., Jr. (2014). Liquidity: Meaning, measurement, management. *Federal Reserve Bank of St. Louis Review, 96*(3), 199-212.

Lucas, R. E., Jr., & Stokey, N. L. (2011). Liquidity crises: Understanding sources and limiting consequences: A theoretical framework. *Federal Reserve Bank of Minneapolis Economic Policy Paper 11-3.*

Lysandrou, P. (2014, May). *Financialisation and the limits of circuit theory*. Paper

presented at the Annual Workshop of the Post Keynesian Economics Study Group, SOAS, University of London. Retrieved from http://postkeynesian.net.

Marsh, T. A. (1984). On Euler-equation restrictions on the temporal behavior of asset returns. *Sloan Working Paper No. 1619-84*.

Martin, F. (2013). *Money: The unauthorised biography*. London, UK: The Bodley Head.

McCloskey, D. N., & Zecher, J. R. ([1976] 1997). How the gold standard worked, 1880-1913. In B. Eichengreen & M. Flandreau (Eds.), *The gold standard in theory and history* (2nd ed.) (pp. 40-53). London, UK: Routledge. (Reprinted from J. A. Frenkel and H. G. Johnson (Eds.) *The monetary approach to the balance of payments*, 1976, London UK: George Allen & Unwin).

McLeay, M., Radia, A., & Thomas, R. (2014). Money creation in the modern economy. *Bank of England Quarterly Bulletin*, Q1.

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, *15*(2), 145-161

Mehrling, P. (1996). The relevance to modern economics of the banking school view. In G. Deleplace & E. J. Nell (Eds.), *Money in motion: The post Keynesian and circulation approaches* (pp. 330-340). New York, NY: St. Martin Press.

Mehrling, P. (1999). The vision of Hyman P. Minsky. *Journal of Economic Behaviour & Organisation*, *39*(2), 129-158.

Mehrling, P. (2000a). Modern money: Fiat or credit? *Journal of Post Keynesian Economics*, *22*(3), 397-406.

Mehrling, P. (2000b, August). *Understanding Fischer Black*. Working paper presented to the NYU Austrian Colloquium.

Mehrling, P. (2000c). *What is monetary economics about?* Barnard College, Columbia University, mimeo.

Mehrling, P. (2010). Monetary policy implementation: A microstructure approach. In R. Leeson (Ed.), *David Laidler's contributions to macroeconomics* (pp. 85-92). Basingstoke, UK: Palgrave Macmillan.

Mehrling, P. (2011). *The new Lombard St: How the Fed become the dealer of last resort*.

Princeton, NJ: Princeton University Press.

Mehrling, P. (2012). Three principles for market-based credit regulation. *American Economic Review*, *102*(3), 107-112.

Mehrling, P. (2013a). The inherent hierarchy of money. In L. Taylor, A. Rezai & T. R. Michl (Eds.), *Social fairness and economics: Economic essays in the spirit of Duncan Foley* (pp. 394-404). Abingdon, UK: Routledge.

Mehrling, P. (2013b). Essential hybridity: A money view of FX. *Journal of Comparative Economics*, *41*(2), 355-363.

Mehrling, P. (2015). Elasticity and discipline in the global swap network. *International Journal of Political Economy*, *44*(4), 311-324.

Mehrling, P. (2016, September 9). Beyond the Taylor rule [Blog post]. Retrieved from http://www.perrymehrling.com/2016/09/beyond-the-taylor-rule/

Mehrling, P., Pozsar, Z., Sweeney, J., & Neilson, D. (2013, November 6). *Bagehot was a shadow banker: Shadow banking, central banking, and the future of global finance*. Retrieved from https://ssrn.com/abstract=2232016

Menger, K. (1892). On the origin of money. *The Economic Journal*, *2*(6), 239-255.

Michell, J. (2017) Do shadow banks create money? 'Financialisation' and the monetary circuit. *Metroeconomica*, *68*(2), 354-377.

Minsky, H. P. (1957). Central banking and money market changes. *The Quarterly Journal of Economics*, *71*(2), 171-187.

Minsky, H. P. (1975). *John Maynard Keynes*. New York, NY: Columbia University Press.

Minsky, H. P. ([1986] 2008). *Stabilizing an unstable economy*. New York, NY: McGraw-Hill.

Mises, L. von (1953). *The theory of money and credit*. Haven, CT: Yale University Press.

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, *48*(3), 261-297.

Moggridge, D. E. (1992). *Maynard Keynes: An economist's biography*. London, UK: Routledge.

Moore, B. J. (1988). *Horizontalists and verticalists: The macroeconomics of credit money*. Cambridge, UK: Cambridge University Press.

Moore, B. J. (1994). The demise of the Keynesian multiplier: A reply to Cottrell. *Journal of Post Keynesian Economics, 17*(1), 121-133.

Moore, B. J. (2001). Some reflections on endogenous money. In L. P. Rochon & M. Vernengo (Eds.), *Credit, interest rates and the open economy: Essays on horizontalism* (pp. 11-30). Cheltenham, UK: Edward Elgar.

Moore, B. J. (2006). *Shaking the invisible hand: Complexity, endogenous money and exogenous interest rates*. Basingstoke, UK: Palgrave Macmillan.

Mott, T. (1985). Towards a post-Keynesian formulation of liquidity preference. *Journal of Post Keynesian Economics, 8*(2), 222-232.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29*(3), 315-335.

Nelson, A. (2005). Marx's objections to credit theories of money. In F. Moseley (Ed.), *Marx's theory of money: Modern appraisals* (pp. 65-77). Basingstoke, UK: Palgrave Macmillan.

Nesvetailova, A. (2007). *Fragile finance: Debt, speculation, and crisis in the age of global credit*. Basingstoke, UK: Palgrave Macmillan.

Nesvetailova, A. (2010). *Financial alchemy in crisis: The great liquidity illusion*. London, UK: Pluto Press.

Niehans, J. (1978). *The theory of money*. Baltimore, MD: John Hopkins University Press.

O'Donnell, R. (2014). A critique of the ergodic/nonergodic approach to uncertainty. *Journal of Post Keynesian Economics, 37*(2), 187-209.

O'Donnell, R. (2016a). Second contribution to the ENE critique: Reply to Davidson, part 1. *Journal of Post Keynesian Economics, 39*(1), 17-43.

O'Donnell, R. (2016b). Third contribution to the ENE critique: Reply to Davidson, part 2. *Journal of Post Keynesian Economics, 39*(2), 145-171.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance, 58*(4), 1335-1354.

Palley, T. I. (2015). The critics of modern money theory (MMT) are right. *Review of Political Economy*, *27*(1), 45-61.

Palley, T. I. (2016). Why ZLB economics and negative interest rate policy (NIRP) are wrong: A theoretical critique. *IMK Working Paper No. 172*.

Parguez, A., & Seccareccia, M. (2000). The credit theory of money: The monetary circuit approach. In J. Smithin (Ed.), *What is money?* (pp. 101-123). London, UK: Routledge.

Passarella, M. V. (2014). Financialization and the monetary circuit: A macro-accounting approach. *Review of Political Economy*, *26*(1), 128-148.

Pastor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642-685.

Patinkin, D. (1965). *Money, interest, and prices: An integration of monetary and value theory* (2nd ed.). New York, NY: Harper & Row.

Peters, O., & Gell-Mann, M. (2016). Evaluating gambles using dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *26*(2), 023103.

Pivetti, M. (2009). Rate of interest. In J.E. King (Ed.), *The Elgar companion to post Keynesian economics* (2nd ed.) (pp. 474-478). Cheltenham, UK: Edward Elgar.

Pozsar, Z. (2014, July 2). *Shadow banking: The money view*. Retrieved from http://ssrn.com/abstract=2476415

Pozsar, Z. (2015, November 5). *A macro view of shadow banking: Levered betas and wholesale funding in the context of secular stagnation*. Retrieved from https://ssrn.com/abstract=2232016

Pozsar, Z. (2016). From exorbitant privilege to existential trilemma. *Credit Suisse Fixed Income Research, Global Money Notes #8.* Retrieved from https://research-doc.credit-suisse.com/docView?language=ENG&format=PDF&sourceid=em&document_id=1067001821&serialid=FdgDLSRBS51YJLr69%2bcO6H1iGqGyLNuzqEDE5DwoUt8%3d

Pozsar, Z. (2017). Sterilization and the fracking of reserves. *Credit Suisse Fixed Income Research, Global Money Notes #10.* Retrieved from https://research-doc.credit-suisse.com/docView?language=ENG&format=PDF&sourceid=csplusresearchcp&document_id=1079774201&serialid=OFfoqqK1y%2B637qpKPp5bxWmkgh3lbdHO1ODUsN0nrv4%3D

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1-2), 122-136.

PwC. (2015). *Global financial markets liquidity study*. Retrieved from http://www.pwc.com/gx/en/financial-services/publications/assets/global-financial-market-liquidity-study.pdf

Quadrini, V. (2011). Financial frictions in macroeconomic fluctuations. *Federal Reserve Band of Richmond Economic Quarterly*, *97*(3), 209-254.

Radcliffe Committee. (1959). *Report of the Committee on the Working of the Monetary System, Cmnd. 827*. London, UK: HMSO.

Radner, R. (1968). Competitive equilibrium under uncertainty. *Econometrica, 36*(1), 31-58.

Rajan, R. (2010). *Fault lines: How hidden fractures still threaten the world economy*. Princeton, NJ: Princeton University Press.

Ram, A. (2016, October 9). Fees for money market funds reinstated. *Financial Times*. Available from https://www.ft.com/content/3cb02528-8ada-11e6-8cb7-e7ada1d123b1

Reserve Bank of Australia. (2016). *Management and Use of an Exchange Settlement Account*. Retrieved June 26, 2016, from http://www.rba.gov.au/payments-and-infrastructure/esa/#management

Ricks, M. (2011). Regulating money creation after the crisis. *Harvard Business Law Review, 1*(1), 75-143.

Robertson, D. H. (1940). *Essays in monetary theory*. London, UK: P. S. King & Son.

Robinson, J. (1951). The rate of interest. *Econometrica*, *19*(2), 92-111.

Rochon, L. P. (1997). Keynes's finance motive: A re-assessment. Credit, liquidity preference and the rate of interest. *Review of Political Economy*, *9*(3), 277-293.

Rochon, L. P. (1999). The creation and circulation of endogenous money: A circuit dynamique approach. *Journal of Economic Issues*, *33*(1), 1-21.

Rochon, L. P., & Rossi, S. (2013). Endogenous money: The evolutionary versus revolutionary views. *Review of Keynesian Economics, 1*(2), 210-229.

Rochon, L. P., & Setterfield, M. (2007). Interest rates, income distribution, and monetary

policy dominance: Post Keynesians and the "fair rate" of interest. *Journal of Post Keynesian Economics, 30*(1), 13-42.

Rogers, C. (1989). *Money, interest and capital: A study in the foundations of monetary theory*. Cambridge, UK: Cambridge University Press.

Rogers, C. (2006a). Exogenous interest rates and modern monetary theory and policy: Moore in perspective. In M. Setterfield (Ed.), *Complexity, endogenous money and macroeconomic theory: Essays in honour of Basil J. Moore* (pp. 290-305). Cheltenham, UK: Edward Elgar.

Rogers, C. (2006b). Cochrane's defence of the fiscal theory of the price level – clarification and critique. *University of Adelaide Research Paper No. 2006-03*.

Rogers, C. (2006c). Doing without money: A critical assessment of Woodford's analysis. *Cambridge Journal of Economics, 30*(2), 293-306.

Rogers, C. (2007). One hundred years from today [Review of Fisher, Irving. The rate of interest: Its nature, determination and the relation to economic phenomena (1907).]. *History of Economics Review, 46*(1), 136-150.

Rogers, C. (2008a). Money as friction: Conceptual dissonance in Woodford's interest and prices. *University of Adelaide Research Paper No. 2008-03.*

Rogers, C. (2008b, March). *Keynes, Keynesians and contemporary monetary theory and policy: An assessment*. Paper presented at the 4th International Conference on Keynes's Influence on Modern Economics - The Keynesian Revolution Reassessed, at Sophia University, Tokyo.

Rogers, C. (2014). *Is monetary theory for real? Why attempts to introduce money into real macroeconomic models fail*. University of Adelaide, mimeo.

Romanchuk, Brian. (2016, July 24). Seriously, money is not a zero coupon perpetual [Blog post]. Retrieved July 29, 2016, from http://www.bondeconomics.com/2016/07/seriously-money-is-not-zero-coupon.html

Rowe, Nick. (2016, July 24). Some simple basic money, for finance people [Blog post]. Retrieved July 29, 2016, from http://worthwhile.typepad.com/worthwhile_canadian_initi/2016/07/simple-basic-money-for-finance-people.html

Rossi, S. (2012). Chartalism. In J. E. King (Ed.), *The Elgar companion to post Keynesian economics* (2nd ed.) (pp. 78-82). Cheltenham, UK: Edward Elgar.

Runde, J. (1990). Keynesian uncertainty and the weight of arguments. *Economics and Philosophy*, *6*(2), 275-292.

Runde, J. (1994). Keynesian uncertainty and liquidity preference. *Cambridge Journal of Economics*, *18*(2), 129-144.

Sachs, J., & Zini, A. A., Jr. (1996). Brazilian Inflation and the Plano Real. *The World Economy*, *19*(1), 13-37.

Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *The Journal of Political Economy*, *66*(6), 467-482.

Samuelson, P. A. ([1980] 2015). A corrected version of Hume's equilibrating mechanisms for international trade. In S. G. Medema & A. M. C. Waterman (Eds.), *Paul Samuelson on the history of economic analysis: Selected essays* (pp. 39-58). New York, NY: Cambridge University Press.

Sargent, T. J., & Wallace, N. (1975). "Rational" expectations, the optimal monetary instrument, and the optimal money supply rule. *Journal of Political Economy*, *83*(2), 241-254.

Sawyer, M. (2013). Endogenous money, circuits and financialization. *Review of Keynesian Economics*, *1*(2), 230-241.

Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). London, UK: Routledge.

Sayers, R. S. (1960). Monetary thought and monetary policy in England. *The Economic Journal*, *70*(280), 710-724.

Sayers, R. S. (1964). *Modern banking* (6th ed.). London, UK: Oxford University Press.

Schumpeter, J. A. (1954). *History of economic analysis*. New York, NY: Oxford University Press.

Schwartz, A. (1973). Secular price change in historical perspective. *Journal of Money, Credit, and Banking*, *5*(1), 243-269.

Seccareccia, M. (2012). Financialization and the transformation of commercial banking:

Understanding the recent Canadian experience before and during the international financial crisis. *Journal of Post Keynesian Economics*, *35*(2), 277-300.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425-442.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, *71*(3), 421-36.

Shin, H. S. (2016, November). *The bank/capital markets nexus goes global*. Speech at the London School of Economics and Political Science, London, UK. Retrieved from http://www.bis.org/speeches/sp161115.pdf

Shin, H. S. (2017, March). *How much should we read into shifts in long-dated yields?* Speech at the US Monetary Policy Forum, New York, NY. Retrieved from http://www.bis.org/speeches/sp170303.pdf

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35-55.

Skidelsky, R. (1992). *John Maynard Keynes: Vol. 2, the economist as saviour, 1920-1937*. London, UK: Macmillan.

Smith, A. ([1776] 1910). *An inquiry into the nature and causes of the wealth of nations* (Everyman's Library ed.). London, UK: J. M. Dent.

Smithin, J. (2006). The theory of interest rates. In P. Arestis & M. Sawyer (Eds.), *A handbook of alternative monetary economics* (pp. 273-290). Cheltenham, UK: Edward Elgar.

Smithin, J. (2017). Interest rate determination and endogenous money. In L. P. Rochon & S. Rossi (Eds.), *Advances in endogenous money analysis* (pp. 262-285). Cheltenham, UK: Edward Elgar.

Sowell, T. (1974). *Classical economics reconsidered*. Princeton, NJ: Princeton University Press.

Stigum, M. L. (1990). *The money market* (3rd ed.). Homewood, IL: Business One Irwin.

Summers, L. H. (1985). On economics and finance. *The Journal of Finance*, *40*(3), 633-635.

Summers, L. H. (2016). Secular stagnation and monetary policy. *Federal Reserve Bank of St. Louis Review*, *98*(2), 93-110.

Sushko, V., Borio, C., McCauley, R., & McGuire, P. (2016). The failure of covered interest parity: FX hedging demand and costly balance sheets. *BIS Working Paper No. 590*.

Taleb, N. N. (2007). Black swans and the domains of statistics. *The American Statistician*, *61*(3), 198-200.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*(1), 195-214.

Tily, G. (2006). Keynes's theory of liquidity preference and his debt management and monetary policies. *Cambridge Journal of Economics*, *30*(5), 657-670.

Tirole, J. (2008). Liquidity shortages: Theoretical underpinnings. *Banque de France Financial Stability Review*, *11*(1), 53-63.

Tirole, J. (2011). Illiquidity and all its friends. *Journal of Economic Literature*, *49*(2), 287-325.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies*, *25*(2), 65-86.

Tobin, J. (1963). Commercial banks as creators of "money". In D. Carson (Ed.), *Banking and monetary studies* (pp. 408-419). Homewood, IL: R. D. Irwin.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit, and Banking*, *1*(1), 15-29.

Tobin, J., & Golub, S. S. (1998). *Money, credit, and capital*. Boston, MA: Irwin/McGraw-Hill.

Townshend, H. (1937). Liquidity-premium and the theory of value. *The Economic Journal*, *47*(185), 157-169.

Treynor, J. L. (1987). The economics of the dealer function. *Financial Analysts Journal*, *43*(6), 27-34.

Turner, A. (2013, September). *Credit, money and leverage: What Wicksell, Hayek and Fisher knew and modern macroeconomics forgot*. Paper presented at the Stockholm

School of Economics Conference Towards a Sustainable Financial System, Stockholm, Sweden.

Tymoigne, E. (2017). On the centrality of redemption: Linking the state and credit theories of money through a financial approach to money. *Levy Economics Institute Working Papers Series No. 890.*

Tymoigne, E., & Wray, L. R. (2006). Money: An alternate story. In P. Arestis & M. Sawyer (Eds.), *A handbook of alternative monetary economics* (pp. 1-16). Cheltenham, UK: Edward Elgar.

Tymoigne, E., & Wray, L. R. (2013). Modern money theory 101: A reply to critics. *Levy Economics Institute Working Papers Series No. 778*.

Vayanos, D., & Wang, J. (2011). Theories of liquidity. *Foundations and Trends in Finance*, *6*(4), 221-317.

Wallace, N. (1981). A Modigliani-Miller theorem for open-market operations. *The American Economic Review*, *71*(3), 267-274.

Weatherson, B. (2002). Keynes, uncertainty and interest rates. *Cambridge Journal of Economics*, *26*(1), 47-62.

Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, *24*(3), 401–421.

Weitzman, M. L. (2007). Subjective expectations and asset-return puzzles. *The American Economic Review*, *97*(4), 1102-1130.

Wells, P. (1983). A post Keynesian view of liquidity preference and the demand for money. *Journal of Post Keynesian Economics*, *5*(4), 523-536.

Werner, R. A. (2015). A lost century in economics: Three theories of banking and the conclusive evidence. *International Review of Financial Analysis*, *46*(1), 361-379.

Wicksell, K. ([1898] 1936). *Interest and prices* (R. F. Kahn, Trans.). London, UK: Macmillan. (Original work published 1898)

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton, NJ: Princeton University Press.

Woodford, M. (2012, September). *Methods of policy accommodation at the interest-rate*

lower bound. Paper presented at the 2012 Jackson Hole Symposium: The Changing Policy Landscape, Jackson Hole, WY.

Wray, L. R. (1991a). Endogenous money and a liquidity preference theory of asset prices. *Review of Radical Political Economics*, *23*(1-2), 118-125.

Wray, L. R. (1991b). Boulding's balloons: A contribution to monetary theory. *Journal of Economic Issues*, *25*(1), 1–20.

Wray, L. R. (1992a). Alternative theories of the rate of interest. *Cambridge Journal of Economics*, *16*(1), 69-89.

Wray, L. R. (1992b). Commercial banks, the central bank, and endogenous money. *Journal of Post Keynesian Economics*, *14*(3), 297–310.

Wray, L. R. (1998). *Understanding modern money: The key to full employment and price stability*. Cheltenham, UK: Edward Elgar.

Wray, L. R. (Ed.) (2004). *Credit and state theories of money: The contributions of A. Mitchell Innes*. Cheltenham, UK: Edward Elgar.

Wray, L. R. (2006). When are interest rates exogenous? In M. Setterfield (Ed.), *Complexity, endogenous money and macroeconomic theory: Essays in honour of Basil J. Moore* (pp. 271-289). Cheltenham, UK: Edward Elgar.

Wray, L. R. (2007). Endogenous money: Structuralist and horizontalist. *Levy Economics Institute Working Papers Series No. 512*.

Young, A. A. ([1924] 1999). The mystery of money: How modern methods of making payments economize the use of money. The role of checks and bank-notes. The enormous edifice of credit. In P. G. Mehrling & R. J. Sandilands (Eds.), *Money and growth: Selected papers of Allyn Abbott Young* (pp. 265-276). London, UK: Routledge. (Reprinted from *The book of popular science*, 1924, New York, NY: The Grolier Society).