Kyoto University-Durham University Joint International Symposium 2012:

Emergence and Feedback in Physical and Social Systems





Period: November 27-29 (Tuesday-Thursday), 2012

Venue: ICR Seminar Hall, Kyoto University
(Uji, Kyoto 611-0011, Japan)

Organized by

Kyoto University (*Institute for Chemical Research*, *International Research Unit of Integrated Complex System Science*, *Institute of Economic Research*, *Yukawa Institute for Theoretical Physics*)

and

Durham University (*Department of Physics*, *Department of Mathematical Sciences*, *Department of Theology and Religion*, *Department of Economics*, *Durham University International Office*)

Supported by

ICR-JURC, the Kyoto University Foundation, and the Daiwa Foundation

Kyoto University-Durham University Joint International Symposium 2012:

Emergence and Feedback in Physical and Social Systems





Period: November 27-29 (Tuesday-Thursday), 2012

Venue: ICR Seminar Hall, Kyoto University
(Uji, Kyoto 611-0011, Japan)

Organized by

Kyoto University (*Institute for Chemical Research*, *International Research Unit of Integrated Complex System Science*, *Institute of Economic Research*, *Yukawa Institute for Theoretical Physics*)

and

Durham University (*Department of Physics*, *Department of Mathematical Sciences*, *Department of Theology and Religion*, *Department of Economics*, *Durham University International Office*)

Supported by

ICR-JURC, the Kyoto University Foundation, and the Daiwa Foundation

Preface:
An Emerging Opportunity

Tom McLeish
Pro-Vice-Chancellor for Research
Durham University

For most of the participants in this symposium, the activity that it has generated will have started with the "opportunity" to think about the long abstracts of their contributions. But for Hiroshi Watanabe, Seth Kunin and me the idea has been working away for much longer. When ten or more independent collaborations of academics between Kyoto University and Durham University choose to work together, it suggests that there might be a natural complementarity between the two institutions that is worth exploring further. You could almost conjecture that this is an emergent property of two universities. So it is proving to be, when the collective of Kyoto-Durham research collaborations is examined.

We thought we would test this conjecture experimentally – and this symposium is the central part of the experiment. Other preliminary results are already known: colleagues have responded with interest and energy to the idea (so far in line with the hypothesis). I am looking forward immensely to hearing more about the topics in the abstracts. Equally encouraging was the enthusiasm of the Daiwa Foundation for this project, and the helpful way in which their staff guided the successful application for funds which has supported most of the long-distance travel it required. This gives us the opportunity to thank the foundation for that support (again, so far so good). The support from the Kyoto University Foundation and ICR-JURC is also acknowledged with thanks. But the project is a very bold one indeed: a radical exercise in interdisciplinary dialogue across the arts, humanities, social sciences and sciences, and between academics spanning different cultures and languages.

I think it was Hiroshi who came up with the idea of a symposium on emergent phenomena. If any theme stands a chance of creating a focus for interdisciplinary and intercultural dialogue it is surely this one. The great high-medieval scholars in Europe began to map out how one works ones way up and down the inductive tree of knowledge, fuelled by a dual intellectual heritage of Hebrew and Greek thought. In Japan, science in its broadest sense emerged from a notion of good governance, as the beautiful shrine of mirrors (Kitano Tenmangu) in Kyoto testifies. Bringing our long intellectual traditions together, as well as our disciplines and their focused questions, is going to be a rich experience. It will need, however, huge patience. We are tackling and doing something to reverse the fragmentation of disciplines in academia, but at the same time we are confronting barriers of languages in different fields.

So let us be very kind to each other, without losing our academic critical edge. And let us also look beyond this symposium to what our two ancient institutions of learning might continue to do together in future.

Program

Nov.27

09:00-09:15 Opening Remarks — Tom McLeish, Seth Kunin (Durham University), Hiroshi Watanabe (Kyoto University)

Session I (Chair: Tom McLeish, Durham University)

09:15-10:05 **01.** *Nonlinear dynamics in economic theory*
Kazuo Nishimura (Institute of Economic Research, Kyoto University) 1

10:05-10:55 **02.** *Assessing the impact of Islamic finance: Between the aspirations of Islamic moral economy and the realities of financial markets*
Mehmet Asutay (School of Government and International Affairs, Durham University) 13

10:55-11:20 **03.** *Multiple equilibria in a dynamic two country model*
Kazumichi Iwasa (Institute of Economic Research, Kyoto University) 15

11:20-12:00 *Free Discussion* (Leader: Kazuo Nishimura, Kyoto University)

Lunch (12:00-13:00)

Session II (Chair: Seth Kunin, Durham University)

13:00-13:50 **04.** *Adiabatic quantum pump*
Hisao Hayakawa (Yukawa Institute for Theoretical Physics, Kyoto University) 24

13:50-14:40 **05.** *The emergence of the chemical bond*
Stewart Clark (Department of Physics, Durham University) 26

14:40-15:05 **06.** *Toward unveiling the explosion mechanism of gamma-ray bursts*
Shigehiro Nagataki (Yukawa Institute for Theoretical Physics, Kyoto University) 34

15:05-15:30 **07.** *Quantum physics from emergence: Materials, experiments and calculations*
Tom Lancaster (Department of Physics, Durham University) 37

15:30-16:00 *Free Discussion* (Leader: Hisao Hayakawa, Kyoto University)

Session III

16:00-17:45 *Poster Presentation (P-01 – P-17) and Collaborative Breakout Discussions*

P-01. *The emergence of "Tipping Points" games in climate change debates: Scientific versus public discoursess*
Pojanath Bhatanacharoen (Durham Business School, Durham University) 45

P-02. *Beauty as protection against environmental instability*
James Woodward (Department of Philosophy, Durham University) 50

P-03. *Emergence of conceptual foundation blocks of chemistry from fundamental quantum mechanics*
Paul Tulip (Department of Physics, Durham University) 55

P-04. *Direct numerical simulation of electrophoresis of charge colloids in AC field*
ChunYu Shih (Department of Chemical Engineering, Kyoto University) 57

P-05. *Hydrodynamic interactions in a compressible fluid*
Rei Tatsumi (Department of Chemical Engineering, Kyoto University) 60

P-06. *Direct numerical simulations of sedimentation using a smooth profile method*
Adnan Hamid (Department of Chemical Engineering, Kyoto University) 62

P-07. *Molecular simulations of polymers under elongational deformations*
Yuichi Masubuchi (Institute for Chemical Research, Kyoto University) 65

P-08. *Thoughts on modelling emergent phenomena*
Sathish Sukumaran (Graduate School of Science and Engineering, Yamagata University) 69

P-09. *Sparse multiple graph integration for label propagation*
Masayuki Karasuyama (Institute for Chemical Research, Kyoto University) 75
P-10. *Integer programming method for completing ortholog-based gene-protein-reaction network by gene essentiality*
Takeyuki Tamura (Institute for Chemical Research, Kyoto University) 77
P-11. *Latent feature models for biological networks*
Canh Hao Nguyen (Institute for Chemical Research, Kyoto University) 80
P-12. *Identifying neighborhoods of coordinated gene expression and metabolite profiles*
Timothy Hancock (Institute for Chemical Research, Kyoto University) 82
P-13. *Cationic state distributions over Chlorophyll pairs in photosystem I and II*
Keisuke Saito (Career-Path Promotion Unit for Young Life Scientists, Kyoto University) 84
P-14. *Protein folding gone wrong, feedback and Parkinson's disease*
Max Brown (School of Biological and Biomedical Sciences, Durham University) 89
P-15. *The subjectivity involved in objectively reporting themes and topics in the media*
Laurens de Rooij (Department of Theology & Religion, Durham University) 93
P-16. *Emerging from the cell: The role of the cytoskeleton in determining cell architecture*
David Mentlak (School of Biological and Biomedical Sciences, Durham University) 101
P-17. *The emerging patterns and problems of climate change: What can we learn from arctic plants?*
Rachael Oakenfull (School of Biological and Biomedical Sciences, Durham University) 106

Nov.28

Session IV (Chair: Hiroshi Mamitsuka, Kyoto University)

09:00-09:50 **08.** *Physicists behaving badly Or: (Why) does physics play a natural role in interdisciplinary studies of emergence?*
Tom McLeish (Department of Physics, Durham University) 111
09:50-10:40 **09.** *Softmatter physics: Multiscale simulations for polymer melt flow*
Takashi Taniguchi (Department of Chemical Engineering, Kyoto University) 118
10:40-11:30 **10.** *Tipping points and emergence in the behaviour of landslides*
David Petley (Department of Geography, Durham University) 128
11:30-11:55 **11.** *Landslides induced by typhoon Talas 2011 and long-term landscape development*
Masahiro Chigira (Disaster Prevention Research Institute, Kyoto University) 130
11:55-12:25 *Free Discussion* (Leader: Hiroshi Watanabe, Kyoto University)

Lunch (12:25-13:30)

Session V (Chair: Hiroshi Watanabe, Kyoto University)

13:30-14:20 **12.** *Literature and/as tipping point. Reflections on literary form, consciousness and emergence in the eco-literature of the Swiss writer Franz Hohler*
Nicholas Saul (School of Modern Language and Cultures, Durham University) 136
14:20-15:10 **13.** *Emergence and reduction: Where is the evidence?*
Robin Hendry (Department of Philosophy, Durham University) 147

15:10-15:35 **14.** *The dynamics of culture change: Gradual or punctuated*
Seth Kunin (Department of Theology and Religion, Durham University) ... 157
15:35-16:30 *Free Discussion/Collaborative breakout discussion*
(Leader: Seth Kunin, Durham University)

18:00-19:30 **Symposium dinner**

Nov.29

Session VI (Chair: Takashi Taniguchi, Kyoto University)

09:00-09:50 **15.** *Mining metabolic pathways through gene expression*
Hiroshi Mamitsuka (Institute for Chemical Research, Kyoto University) 169
09:50-10:40 **16.** *How shape and function of cells and tissues is determined by intermediate filaments and their associated protein chaperones*
Roy Quinlan (School of Biological and Biomedical Sciences, Durham University) 179
10:40-11:05 **17.** *Chemistry of low-barrier hydrogen bonds in protein active sites*
Hiroshi Ishikita (Career-Path Promotion Unit for Young Life Scientists, Kyoto University) 183
11:05-11:30 **18.** *Disulfide bond formation networks and chaperone interactions that are subject to feedback regulation*
Adam Benham (School of Biological and Biomedical Sciences, Durham University) 200
11:30-12:00 *Free Discussion* (Leader: Hiroshi Mamitsuka, Kyoto University)

Lunch (12:00-13:00)

Session VII (Chair: Nicholas Saul, Durham University)

13:00-13:25 **19.** *Regulation of redox homeostasis and proteostasis in the ER*
Kazuhiro Nagata (Department of Molecular and Cellular Biology, Kyoto Sangyo University) 207
13:25-13:50 **20.** *Application of quantum mechanical formulation to classical stochastic processes*
Jun Ohkubo (Graduate School of Informatics, Kyoto University) 209
13:50-14:15 **21.** *Emergent relations: meanings and materiality in human-environmental interaction*
Veronica Strang (Institute of Advanced Study, Durham University) 215
14:15-14:40 **22.** *Agent-based simulation of emergence in populations of financial organizations*
Philip Garnett (Centre for Coevolution for Biology and Culture, Durham University) 230
14:40-15:05 **23.** *Qatar and Japan: A marriage of convenience?*
David Roberts (School of Government and International Affairs, Durham University) 235

Break (15:05-15:30)

Session VIII

15:30-16:30 *Overview Discussion* (Leader: Tom McLeish, Durham University)
16:30-16:40 *Closing remarks* (Seth Kunin, Durham University)

Nov.27 (oral presentation)

01

Nonlinear Dynamics in Economic Theory

Kazuo Nishimura
Institute of Economic Research,
Kyoto University

Summary: The paper exposits a number of key dynamic models from the field of economics. The models share the feature that evolution of the system is consistent with the optimizing behavior of rational economic agents. Specifically, the laws of motion that we investigate are solutions to discrete time dynamic decision processes. Even with rational agents and simple economic environments, a variety of complex behaviors are shown to obtain.

1. Introduction

The primary concern of economics as a science is allocation of scarce resources among several alternative and competing uses. In this connection, it is important to bear in mind that resources must be allocated not only contemporaneously, but also across time. The planning involved occurs in each period after observation of the current state. Therefore decisions of intertemporal allocation are naturally described by discrete-time dynamical functions.

These dynamical processes are not defined by arbitrarily chosen laws of motion. Agents' decisions are made according to such concerns as profit maximization by firms, utility maximization by households and social welfare maximization by the policy maker. In formulating these plans, forward looking agents consider the affect of their actions on the time path of state variables. The dynamical systems used in economic theory are thus obtained as solutions to such intertemporal optimization problems.

The foundational models in this area are those that replicate growth through the accumulation of productive capital. In economics, interest in the theory of growth has been revived during the last fifteen years. The reason for the resurrection of growth research is that the framework of intertemporal optimization has been found to explain a much wider range of phenomena than was previously believed. Using results from the field of non-linear dynamical systems, it has been shown that intertemporal optimization theory can provide new explanations for business cycles and for international differences in growth and development.

In the economic literature, historically there have been two types of explanations for observed fluctuations in the level of economic activity. One type of explanation has been built on the view that fluctuations are caused by factors that are exogenous to economic systems. For example, agricultural production can be affected by weather, consumers tastes may be influenced by intangible fads, or government policies may change erratically. In such environments the market plays the role of a filter,

passing random shocks to economic variables, which constantly deviate from their trends. These deviations are regarded as business cycles; the field of research is referred to as the "real business cycle" literature.

The other type of explanation views fluctuations as phenomena endogenous to economic systems. This view dominated the literature prior to the 1960s.[1] Those studies in turn were based on the Keynesian premise that certain regularities exist in the relationships between major macroeconomic variables. It has become clear, however, that such a premise is inconsistent with the rational and well-informed behavior of economic agents, and this inconsistency has contributed to a shift of research focus towards what is now known as neoclassical economics. The latter is built on the foundations of rational, optimizing behavior.

The purpose of the neoclassical methodology is not of course to deny the occurrence of irrational behavior by actors in the real economy. The force of its argument comes from its logical consistency and the potential for prediction. Without a priori restrictions on human market and strategic behavior, models consistent with any outcome can be constructed. Theories that cannot preclude any possibility are not useful for structuring academic debate.

If the neoclassical paradigm is accepted, however, it remains unclear whether or not the fundamental structure of an economy itself — without the influence of external noise — may explain business fluctuations in the neoclassical framework. Work on non-linear dynamics from the late 1970s and early 1980s appears to have answered this question in the affirmative, renewing widespread interest in the endogenous explanation of economic fluctuations.[2]

In the first part of this paper, a number of deterministic discrete time models are presented. All treat by intertemporal optimization the fundamental problem of savings and consumption. Some models are associated with asymptotically stable dynamics, while others exhibit cycles and other complex behavior.

In addition, some basic results for stochastic dynamic models are also discussed. Stochastic models are important because of their obvious connections with the empirical literature. In addition, they form the foundations of the so-called real business cycles mentioned above, which remain an active research area. Among the results listed here, particular emphasis is placed on the implications of monotone decision rules for asymptotic stochastic behavior.

---

1 See, for example, Harrod [8], Samuelson [19], or Kaldor [9].

2 Benhabib and Nishimura [2] demonstrate the appearance of periodic cycles in along optimal paths of capital accumulation in a model where agents are fully rational and perfectly informed. Benhabib and Day [1], and Grandmont [7] observe the possibilities of chaotic dynamics in models in which agents are rational and informed in certain limited manners. Boldrin and Montrucchio [3], Deneckere and Pelikan [6], and Nishimura and Yano [16] demonstrate the existence of a chaotic optimal path of capital accumulation in a model with fully rational and perfectly informed agents.

2. Aggregative models of economic growth

2.1 The basic framework

The basic premise of the aggregative model can be described as follows: in each period t, a single homogeneous output, Y_t, is produced from the two homogeneous input factors labor, L_t, and capital, K_t. The technically efficient possibilities for production are summarized by an aggregate production function $F(K_t, L_t)$ which exhibits constant returns to scale, positive marginal productivity, and decreasing marginal rate of substitution.

Constant returns — scalar multiplying the vector of inputs multiplies output by the same amount — is motivated by a replication argument. Diminishing returns to individual inputs when others are held fixed seems a natural assumption, and generates convexities that are central to both optimization and dynamic properties.

Because of constant returns to scale, the output-labor ratio $y_t = Y_t / L_t$ is given by

$$y_t = f(k_t) \tag{1}$$

where $k_t = K_t / L_t$ is the capital-labor ratio and $f(k)=F(k, 1)$. Regarding f, it is assumed that

Assumption 1 *$f:[0, \infty)\rightarrow R$ is continuous everywhere, and twice continuously differentiable on $(0,\infty)$.*

Further,

Assumption 2 *$f(0) = 0, f' > 0, f'' < 0, \lim_{k \downarrow 0} f'(k) = \infty, \lim_{k \uparrow 0} f'(k) = 0.$*

The labor force is assumed to be constant, and capital stock depreciates at positive rate δ. Per capita output may be allocated between consumption and gross investment. Denoting per capita consumption by c_t this implies

$$y_t = c_t + k_{t+1} - (1 - \delta)\, k_t. \tag{2}$$

The initial per capita capital stock k_0 is historically given. Social welfare over the infinite planning period is presumed to be represented by the functional

$$\sum_{t=0}^{\infty} \rho^t u(c_t), \tag{3}$$

where $\rho \in (0, 1)$ is the discount factor. Thus, social welfare is the discounted sum of period-wise utility of per capita consumption.

Assumption 3 *$u:[0, \infty) \to \mathbb{R}$ is continuous, increasing, and twice continuously differentiable on $(0,\infty)$.*

Assumption 4 *On $(0, \infty)$, $u' > 0$ and $u'' < 0$. Also, $\lim_{c \downarrow 0} u'(c) = \infty$ and $\lim_{c \uparrow \infty} u'(c) = 0$.*

A sequence of stocks $(k_t)_{t=0}^{\infty}$ is called a feasible path from k_0 if it satisfies the condition $0 \leq k_{t+1} \leq f(k_t) + (1 - \delta)k_t$ for all $t \geq 0$. For each feasible path there is a corresponding sequence of consumption rates $(c_t)_{t=0}^{\infty}$ determined by (1) and (2). A feasible path is called an interior path if $c_t > 0$ holds for all $t \geq 0$, and it is called stationary if $k_t = k$ for all $t \geq 0$ and some constant $k \geq 0$. An optimal path from k_0 is a feasible path from k_0 that maximizes the objective function (3). By the strict concavity of the utility and production functions we can show that the optimal solution from a given k_0 is unique.

An interior path will be called an Euler path if it satisfies the discrete-time Euler equation

$$u'(c_{t-1}) - \rho u'(c_t)\,[f'(k_t) + 1 - \delta] = 0. \tag{4}$$

If an interior path is optimal then it must be an Euler path. By substituting (1) and (2) into (4), the Euler equation becomes a second order difference equation; given k_0 there are infinitely many paths that satisfy it. To distinguish the unique optimum path from the other Euler paths we need an additional optimality condition. In fact it is known that under the assumptions stated above, any Euler path which satisfies the transversality condition

$$\lim_{t \to \infty} \rho^t u'(c_t)\,[f'(k_t) + 1 - \delta]\, k_t = 0$$

is an optimal path (See McKenzie [12]).

2.2 Main results

Because of Assumption 4, every optimal path from a positive initial stock $k_0 > 0$ is an interior path and, consequently, it must be an Euler path. A stationary optimal path $(k, k, k, \ldots)$ with $k > 0$ must satisfy the Euler equation (4) and, hence,

$$\rho^{-1} = f'(k) + 1 - \delta. \tag{5}$$

A solution to (5) is called a steady state.

Local behavior of the solutions around the steady state may be determined by the characteristic equation

$$\lambda^2 \rho u''(c^*) - \lambda[(1+\rho)u''(c^*) + \rho u'(c^*) f''(k^*)] + u''(c^*) = 0$$

evaluated at that point. This polynomial equation has two roots, the product of which is $\rho^{-1} > 1$. Evidently it can never have two roots inside the unit circle. Also, the left-hand side is equal to $u''(c^*) < 0$ when $\lambda = 0$ and equal to $-\rho u'(c^*) f''(k^*) > 0$ when $\lambda = 1$. Evidently there is always one positive root inside the unit circle. This implies that the steady state is locally a saddle-point. The root inside the unit circle corresponds to the optimal solution, because the path converging to the steady state satisfies the transversality condition. Since that root is positive convergence must be monotone.

The above argument is limited to local dynamics of solutions of the Euler equation (4). However the following global result from the maximizing problem (1)–(3) can be proved. It is the discrete-time version of a result originally due to Cass [4] and Koopmans [10].

Theorem 1 *Consider the model defined by (1)–(3). Under Assumptions 1–4 there exists a unique steady state k^*. Moreover, any optimal path from $k_0 > 0$ is monotone and converges to k^*.*

3. Two-sector models

3.1 The basic framework

In this section the aggregative model of Section 2 is extended to the two-sector model. There are two goods: a pure consumption good, C, and a pure capital good, K. Each sector uses both capital and labor as inputs. Capital input must be made one period prior to the production of output. Labor input is made in the same period as output is produced. Denote by $F_C(K_C, L_C)$ and $F_K(K_K, L_K)$ the production functions of sectors C and K, respectively, where K_i and L_i denote the factor inputs in sector $i \in \{C, K\}$. The production functions are assumed to be increasing in each argument, concave and homogeneous of degree one.

The labor endowment of the economy is constant. Without loss of generality we normalize it to *1*. Denote by c_t and y_t the time t per capita outputs of sectors C and K, respectively. Thus we have

$$c_t = F_C\,(K_{C,t-1}, L_{C,t}), \tag{6}$$

$$y_t = F_K\,(K_{K,t-1}, L_{K,t}). \tag{7}$$

Moreover, denote by k_{t-1} the aggregate capital input:

$$K_{C,t} + K_{K,t} = k_t. \tag{8}$$

The output of the capital good sector, y_t, represents the gross accumulation of capital;

$$y_t = k_t - (1 - \delta)\, k_{t-1}, \tag{9}$$

where $\delta \in (0,\ 1)$ is the rate of depreciation. Since the total labor force in the economy has been normalized to *1*,

$$L_{C,t} + L_{K,t} = 1. \tag{10}$$

As before, $u(c)$ is the representative consumer's contemporaneous utility when he consumes c units of the consumption good. With these notations, the two-sector optimal growth model is described by the maximization problem

$$\begin{aligned} &\text{maximize } \textstyle\sum_{t=1}^{\infty} \rho^t u(c_t) \\ &\text{subject to } k_0 = \bar{k}_0 \text{ and constraints (6)-(10),} \end{aligned} \tag{11}$$

where as before $\rho \in (0,\ 1)$ is the discount factor.

In order to analyze the dynamics of the above model it is convenient to express, for each given amount of capital input k, the trade-off between the two outputs by $c = T\,(k, y)$. That is,

$$T\,(k, y) = \max F_C(K_C, L_C) \quad \text{subject to} \begin{cases} F_K(K_K, L_K) = y, \\ L_C + L_K = 1, \\ K_C + K_K = k. \end{cases} \tag{12}$$

The domain of the function T is $\Omega := \{(k, y) \mid k \geq 0,\ 0 \leq y \leq F_K(k, 1)\}$. With this definition, the optimal growth model (11) can be transformed as follows:

$$\text{maximize } \sum_{t=1}^{\infty} \rho^t U(k_{t-1}, k_t)$$
$$\text{subject to } k_0 = \bar{k}_0 \text{ and } 0 \leq k_t \leq F_K(k_{t-1}, 1) + (1-\delta)\, k_{t-1},$$

where $U(x, z) := u(T(x, z - (1-\delta)\, x))$ is called a reduced form utility function.

3.2 Optimal cycles

In this section we assume that the period-wise utility function is linear [i.e., $u(c) = c$] and that capital fully depreciates within one period ($\delta = 1$). The reduced form utility function is then identical to the social production function. That is, $U(k, y) = T(k, y)$. Even if the utility function is linear, it can be shown that the optimal path is still unique under plausible conditions on the production functions.

The Euler equation in the two-sector optimal growth model is

$$U_2(k_{t-1}, k_t) + \rho U_1(k_t, k_{t+1}) = 0, \tag{13}$$

where $U_1(k, y) = \partial U(k, y) / \partial k$ and $U_2(k, y) = \partial U(k, y) / \partial y$. Any path satisfying the Euler equation and the transversality condition

$$\lim_{t \to \infty} \rho^t k_t\, U_1(k_t, k_{t+1}) = 0$$

is known to be optimal.

The steady state k^* corresponds to a stationary solution $(k^*, k^*, k^*, \ldots)$ of (13). The local behavior around k^* is determined by the roots of the characteristic equation

$$\rho U_{12}(k^*, k^*)\, \lambda^2 + [\rho U_{11}(k^*, k^*) + U_{22}(k^*, k^*)]\lambda + U_{21}(k^*, k^*) = 0. \tag{14}$$

evaluated at the steady state.

As in the aggregated model, the product of the two roots is equal to $\rho^{-1} > 1$. Therefore at least one root is outside the unit circle. However, unlike the aggregated model, the other root of (14) is not necessarily inside the unit circle. If ρ is sufficiently small then both roots may be outside the unit circle. On the other hand, since there exists always a unique optimal path starting from k_0, there must still be

a path that satisfies the transversality condition. We will characterize its behavior below.

Consider the case in which the production functions in both sectors have the so-called Cobb-Douglas form

$$F_C(K_C, L_C) = K_C^{\alpha} L_C^{1-\alpha}, \quad 0 < \alpha < 1; \tag{15}$$

$$F_K(K_K, L_K) = K_K^{\beta} L_K^{1-\beta}, \quad 0 < \beta < 1. \tag{16}$$

From the first order conditions of (12), we have $(K_K, L_K)/(K_C, L_C) = [\beta/(1-\beta)]/[\alpha/(1-\alpha)]$. Hence

$$(K_K/L_K) - (K_C/L_C) \begin{cases} > 0 & if \ \beta > \alpha, \\ < 0 & if \ \beta < \alpha. \end{cases} \tag{17}$$

Here K_i/L_i is called the *factor intensity* of sector i, and the left-hand side of (17) is called the factor intensity difference. If $\beta > \alpha$, the production of consumption goods is more labor intensive than the production of capital goods. If $\beta < \alpha$, the converse is true.

In the two sector model with Cobb-Douglas production functions and linear utility, the sign of the cross partial derivative $U_{12}(x, y)$ is determined by the factor intensity difference of the consumption good sector and the capital good sector (Benhabib and Nishimura [2]). This fact, together with the relation (17), implies that

$$U_{12}(x, y) \begin{cases} > 0 & if \ \beta > \alpha, \\ < 0 & if \ \beta < \alpha. \end{cases} \tag{18}$$

We know that given an initial capital stock, there exists a unique optimal path. Therefore optimal paths in this model must be described by a difference equation of the form $k_{t+1} = h(k_t)$. The function h is called the optimal policy function.

Benhabib and Nishimura [2] have shown that the sign of the cross partials of the reduced form utility function determines whether $h(k_t)$ is increasing or decreasing. This, together with equation (18), implies that, in the case $\beta > \alpha$, the graph of the optimal policy function h is strictly increasing whenever it lies in the interior of Ω. Analogously, if $\alpha > \beta$, then the graph of h is strictly decreasing on the interior of Ω. In the case of $\alpha > \beta$, the optimal policy function h becomes a unimodal map, because the graph of h increases along the boundary of Ω and decreases in the interior of Ω.[3]

---

[3] A function $h : [a, b] \to [a, b]$ is called unimodal if there exists $\bar{x} \in [a, b]$ such that h is strictly increasing on $[a, \bar{x}]$ and strictly decreasing on $[\bar{x}, b]$.

The global dynamics of the two sector model with Cobb-Douglas production technology and total capital depreciation ($\delta = 1$) is studied in Nishimura and Yano [17]. For this economy the steady state value is

$$k^* = \left[\frac{\alpha(1-\beta)}{\beta[(1-\alpha)+\rho\delta(\alpha-\beta)]}\right]^{(\rho\beta)^{1/(1-\beta)}}, \tag{19}$$

and the roots of (14) are

$$\lambda_1 = \frac{\beta-\alpha}{1-\alpha}, \quad \lambda_2 = \frac{1-\alpha}{\rho(\beta-\alpha)}. \tag{20}$$

If $\alpha > \beta$, then both roots are negative. The following theorem of Nishimura and Yano [17] gives conditions for the instability of the steady state and thus for the existence of an optimal cycle of period 2:

Theorem 2 *For the economy described by the production functions (15) and (16), linear utility function u (c) = c, and δ = 1, the following is true. If α > β and also*

$$\rho < \frac{1-\alpha}{(\alpha-\beta)} < 1,$$

then the steady state k^* given in (19) is totally unstable, and there exists an optimal path which is periodic of period 2.

4. Optimal chaos

Boldrin and Montrucchio [3] and Deneckere and Pelikan [6] have provided a constructive method to give examples of two-sector optimal growth models in which the optimal policy function is given by the logistic function $h(x) = 4x\ (1 - x)$. The latter is characterized by chaotic paths. While these results are interesting, however, the examples also show that such optimal chaos requires a sufficiently small discount factor. It is therefore quite natural to ask if there is a general relation between the size of the discount factor and the dynamic complexity of optimal paths. Initial results in this direction

were derived by Sorger [21, 22]. Subsequently, Sorger [23], Mitra [13], Nishimura and Yano [18], and Mitra and Sorger [14, 15] have shown that there exists a sharp upper bound on the set of discount factors that are compatible with chaotic optimal paths.

4.1 Basic concepts

Three possible definitions of complicated dynamics in systems of the form $x_{t+1} = h(x_t)$ are first discussed. Here, x_t is the state of the economy at time t (for example the capital-labor ratio) and h:[*0*, *1*] $\mapsto$ [*0*, *1*] is a continuous function which encodes dynamic properties, such as technology and market structure.

We say that the dynamical system $x_{t+1} = h(x_t)$ exhibits ergodic chaos if there exists an absolutely continuous probability measure μ on the interval [*0*, *1*] which is invariant and ergodic under h. Here, absolutely continuity means existence of Radon-Nikodym derivative with respect to the Lebesgue measure. Invariance of μ under h means that $\mu\{x \in [0, 1] \mid h(x) \in B\} = \mu B$ for all measurable $B \subseteq [0, 1]$. An invariant measure μ is said to be ergodic if, in addition, for every measurable set $B \subseteq [0, 1]$ which satisfies $\{x \in [0, 1] \mid h(x) \in B\} = B$ we have either $\mu B = 0$ or $\mu B = 1$.

We say that the dynamical system $x_{t+1} = h(x_t)$ exhibits geometric sensitivity if there is a real constant $\gamma > 0$ such that the following is true: for any $\tau = 0, 1, 2, \ldots$ there exists $\varepsilon > 0$ such that for all $x, y \in [0, 1]$ with $|x - y| < \varepsilon$ and for all $t \in \{0, 1, \ldots, \tau\}$ it holds that

$$|h^{(t)}(x) - h^{(t)}(y)| \geq (1 + \gamma)^{(t)} |x - y|.$$

Geometric sensitivity implies that small perturbations of the initial conditions are magnified at a geometric rate over arbitrary but finite time periods. Of course, the geometric magnification cannot last indefinitely because the state space [*0*, *1*] of the dynamical system is bounded. Note also that geometric sensitivity implies that there is no stable periodic path of the dynamical system.

Finally, we say that the dynamical system $x_{t+1} = h(x_t)$ exhibits topological chaos if there exists a p-periodic solution for all sufficiently large integers p and there exists an uncountable invariant set $S \subseteq [0, 1]$ containing no periodic points such that

$$\liminf_{t \to \infty} |h^{(t)}(x) - h^{(t)}(y)| = 0 < \limsup_{t \to \infty} |h^{(t)}(x) - h^{(t)}(z)|$$

holds, whenever $x \in S$, $y \in S$, and either $x \neq z \in S$ or z is a periodic point. The set S is called a

scrambled set. The condition displayed above says that any two trajectories starting in the scrambled set move arbitrarily close to each other but do not converge to each other or to any periodic orbit.[4]

References

[1] Benhabib, J. and R. H. Day (1982): "A Characterization of Erratic Trajectories in the Overlapping Generations Model," *Journal of Economic Dynamics and Control 4*, 37–55.

[2] Benhabib, J. and K. Nishimura (1985): "Competitive Equilibrium Cycles," *Journal of Economic Theory 35*, 284–306.

[3] Boldrin, M. and L. Montrucchio (1986): "On the Indeterminacy of Capital Accumulation Paths," *Journal of Economic Theory 40*, 26–39.

[4] Cass, D. (1965): "Optimum Growth in an Aggregative Model of Capital Accumulation," *Review of Economic Studies 32*, 233–240.

[5] Devaney, R. (1989): An Introduction to Chaotic Dynamical Systems, Second edition, Addison-Wesley.

[6] Deneckere, R. and S. Pelikan (1986): "Competitive Chaos," *Journal of Economic Theory 40*, 13–25.

[7] Grandmont, J. (1985): "On Endogenous Competitive Business Cycles," *Econometrica 53*, 995–1045.

[8] Harrod, R. (1936): The Trade Cycle: An Essay, Oxford University Press, Oxford.

[9] Kaldor, N. (1940): "A Model of the Trade Cycles," *Economic Journal 50*, 78–92.

[10] Koopmans, T. C. (1965): "On the Concept of Optimal Growth," in The Econometric Approach to

[11] Development Planning, North-Holland, Amsterdam.

[12] Li, T. Y. and J. Yorke (1975): "Period Three Implies Chaos," *American Mathematical Monthly 82*, 985–992.

[13] McKenzie, L. (1986): "Optimal Economic Growth, Turnpike Theorems and Comparative Dynamics," in Handbook of Mathematical Economics: Vol III (Arrow, K. and M. Intriligator, eds.), North-Holland, Amsterdam, 1281–1355.

[14] Mitra, T. (1996): "An Exact Discount Factor Restriction for Period-Three Cycles in Dynamic Optimization Models," *Journal of Economic Theory 69*, 281–305.

[15] Mitra, T. and G. Sorger (1999a): "On the Existence of Chaotic Policy Functions in Dynamic Optimization," *Japanese Economic Review 50*, 470–484.

[16] Mitra, T. and G. Sorger (1999b): "Rationalizing Policy Functions by Dynamic Optimization,"

[4] See Li and Yorke [11], Sarkovskii [20] and Devaney [5] for additional details of results on chaos.

Econometrica 67, 375–392.

[17] Nishimura, K. and M. Yano (1995b): "Non-Linearity and Business Cycles in a Two-Sector Equilibrium Model: An Example with Cobb-Douglas Production Functions," in *Nonlinear and Convex Analysis in Economic Theory* (Maruyama, T., and W. Takahashi, eds.), Springer-Verlag, Berlin, 231–245.

[18] Nishimura, K. and M. Yano (1996a): "On the Least Upper Bound of Discount Factors that are Compatible with Optimal Period-Three Cycles," *Journal of Economic Theory 69*, 206–333.

[19] Samuelson, P. (1939): "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Studies 21*, 75–78.

[20] Sarkovskii, A. N. (1964): "Coexistence of Cycles of a Continuous Map of a Line into Itself," *Ukrainian Mathematical Journal 16*, 61–71.

[21] Sorger, G. (1992a): *Minimum Impatience Theorems for Recursive Economic Models*, Springer-Verlag, Heidelberg.

[22] Sorger, G., (1992b): "On the Minimum Rate of Impatience for Complicated Optimal Growth Paths," *Journal of Economic Theory 56*, 160–179.

[23] Sorger, G. (1994): "Period Three Implies Heavy Discounting," *Mathematics of Operations Research 19*, 1007–1022.

02

Assessing the Impact of Islamic Finance: Between the Aspirations of Islamic Moral Economy *and* the Realities of Financial Markets

Mehmet Asutay
Reader in Middle Eastern and
Islamic Political Economy & Finance
School of Government and International Affairs
Durham University

Abstract:

Developments in the Islamic finance industry in the last two decades have significant impact on both economies of Muslim geography and global financial markets. Today, Islamic finance industry has come a long way, and established itself as an important part of the global financial system. Double digit annual growth during the last two decades and the potential size of the market guarantees that this development and trend will continue, if not increase substantially for the foreseeable future. Inevitably, this strong performance of Islamic finance should be promising for achieving the goals and the aspirations of the Islamic political economy.

Theoretically, this strong development of Islamic finance should have supported the broader economies of Muslim world to achieve the original aspirations of the founding fathers of Islamic economic thought; which were, but not exclusively limited to, entrepreneurship, trade and economic livelihood, all of which are to support social and economic development, stability, justice and prosperity within the religious, ethical and moral codes of the teachings of Islam. This approach implies that the growth of Islamic finance should be supporting the economies of Muslim world to achieve these original aspirations in two different angles: the financial angle and the financing angle.

For the financial angle, theoretically the development of Islamic finance should have added new and novel veins supporting the financial markets through investors which were sidelined before because of their religious concerns. This increase, by providing extra liquidity, extends the potential pool supporting the system. As a result, the inclusion of Islamic finance should help an economy by making its financial system more robust, thus safeguarding the wealth. Inevitably, this additional liquidity support also points to the expansion in the ability of financial system to spread its risks. Hence, through inclusion of new investors and the expansion in the risk bearing ability of the financial system, theoretically the introduction of Islamic finance should improve the economic stability, as well.

The second influence of Islamic finance on the global financial system comes from the financing side. By the expansion in the reach of Islamic finance, investors and entrepreneurs with solid business ideas, but also with the moral and ethical concerns should gain an opportunity to provide and maintain a financing resource for their businesses. As a result, in theory, Islamic finance through its banking system creates much greater access to

financial resources for entrepreneurs, firms and households, who were previously sidelined in the financial system because of their religious, moral and ethical concerns. This, inevitably, is expected to stimulate growth not only in the financial sector, but also in the real economy and industry.

The last two decades have witnessed the emergence and growth of regional financial centres that focus on Shari'ah compliant financing and investment. In the mist of the global financial crisis, however, it cannot be denied that Islamic financial institutions and economies of these countries have also suffered significant blows, raising questions that Islamic finance might not be immune to the weaknesses of the conventional financial system.

As a reflection of the argument above, this paper aims to address the implications of Islamic finance for Muslim economies by investigating the sustainability of the theoretical benefits introduced above from both financial and financing sides. In the first part, we address the equity, liquidity and stability aspects of Islamic finance's contribution to the financial markets of Muslim economies. Second, we analyze the effects of Islamic banking and finance development on the real sector and broader economic context by examining the contribution of Islamic finance to economic growth and industrial output. Finally, the possible improvement areas under the light of the experiences from the last economic crisis will be discussed for further development.

Keywords: Islamic finance, Islamic banking, stock markets, volatility, economic growth.

03

Multiple equilibria in a dynamic two country model

Kazumichi Iwasa
Institute of Economic Research, Kyoto University

Summary: We consider a dynamic two country model of trade with technological differences across countries and show the existence of multiple equilibrium paths. We will prove that if a consumable capital is more labor intensive than a pure consumption good, multiple equilibrium paths can arise even under the normality assumption in consumption.

1. Introduction

It is widely recognized that multiple equilibrium paths, which we call indeterminacy, can arise in dynamic general equilibrium models when there exist some market distortional factors such as factor-generated externalities, imperfect competition, and so on (e.g. Benhabib, Meng, and Nishimura (2000)).

Nishimura and Shimomura (2006) demonstrated that without such market distortional factors, indeteminacy is still possible in a dynamic two country model of trade with two factors of production, a primary factor (labor) and a reproducible one (capital), based on the standard assumptions in trade theory such that there is neither international lending and borrowing nor factor mobility among countries. As shown in Nishimura and Shimomura (2002) and Shimomura (2007), indeterminacy is closely related to important issues in trade theory such as long-run trade patterns and transfer paradox. Although there are many studies on indeterminacy in two country models, either market distortional factors or inferiority in consumption (one good is an inferior good at the steady state) is assumed to make indeterminacy possible.

In this study, we consider the discrete time version of a dynamic two country model with technological differences across countries of the type proposed by Shimomura (2007), extend it, and show that indeterminacy can arise without any market distortional factor and inferiority in consumption.[1] Thus, indeterminacy proves to be a more commonplace phenomenon in an open economy than previously thought.

2. The model

There are two countries, North and South, and two tradable goods, good 1 and good 2. Good 1 is a pure consumption good, while good 2 is a consumable capital which serves as the numeraire. We assume that while newly produced good 2 is tradable, the existing capital is internationally immobile. Moreover, we assume away international lending and borrowing.

We assume that production technologies available in North and South are internationally different.

[1]Bond and Driskill (2008) demonstrate its possibility by simulating a two country model with a durable good.

In North, both goods can be produced using labor and capital which is reproducible. The technologies are described by increasing, quasi-concave and linearly homogeneous production functions. On the other hand, South can produce good 1 alone using labor only. The supply of labor in each country is constant over time.

2.1. Households

The representative household in each country is assumed to maximize the discounted sum of its utilities derived from consuming both goods. Let E be the period-t expenditure function for the North representative household;[2] we define it as

$$E(p_t, u_t) \equiv \min_{c_{1t}, c_{2t}} p_t c_{1t} + c_{2t} \quad \text{subject to } u_t \le u(c_{1t}, c_{2t}), \tag{1}$$

where p_t, u_t, and c_{it} $(i = 1, 2)$ are the price of good 1, the level of utility, and consumption of good i at period-t, respectively, and we assume that the North household's utility function u is strictly concave with $u_{11} < 0$ and $H \equiv u_{11}u_{22} - (u_{12})^2 > 0$. Notice that E denotes the minimum expenditure to attain the level of utility u_t for a given price p_t. From the envelope properties of the expenditure function we obtain the optimal solutions to this minimization problem as

$$c_{1t} = E_p(p_t, u_t) \quad \text{and} \quad c_{2t} = E(p_t, u_t) - p_t E_p(p_t, u_t). \tag{2}$$

Utilizing the expenditure function, the objective of the North household can be written as follows:

$$\max_{\{u_t\}} \sum_{t=0}^{\infty} u_t \rho^t \quad \text{subject to}$$

$$k_{t+1} = (1 + r_t - \delta)k_t + w_t l - E(p_t, u_t), \quad k_0; \text{ given}, \tag{3}$$

where k_t, r_t, w_t, ρ, δ, and l are capital stock, the rental on capital, the wage rate, the discount rate, the rate of depreciation on capital, and an endowment of labor, respectively, and $\rho, \delta \in (0,1)$.

Associated with the problem, (3), is the Lagrangian

$$\mathcal{L} = \sum_{t=0}^{\infty} \{u_t \rho^t + \mu_t [(1 + r_t - \delta)k_t + w_t l - E(p_t, u_t) - k_{t+1}]\}.$$

The first order conditions are

$$\frac{\partial \mathcal{L}}{\partial u_t} = \rho^t - \mu_t E_u(p_t, u_t) = 0, \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial k_t} = -\mu_{t-1} + \mu_t (1 + r_t - \delta) = 0. \tag{5}$$

Defining $\lambda_t \equiv \mu_t / \rho^t$, we obtain from (4) and (5)

[2]The corresponding behavioral relations for the other (South) country will be denoted by a $*$.

$$0 = 1 - \frac{\lambda_{t-1} E_u(p_t, u_t)}{\rho(1 + r_t - \delta)}, \tag{6}$$

$$0 = -\lambda_{t-1} + \rho\lambda_t(1 + r_t - \delta). \tag{7}$$

For given time profiles $\{p_t\}_{t=0}^{\infty}$, $\{r_t\}_{t=0}^{\infty}$, and $\{w_t\}_{t=0}^{\infty}$, and the initial condition k_0, if a value of λ_{-1} is chosen, the dynamical system which consists of (3), (6), and (7) determines the pair of time profiles $\{k_t\}_{t=0}^{\infty}$ and $\{u_t\}_{t=0}^{\infty}$. The following Lemma states under what conditions the pair is optimal from the viewpoint of the North household.

Lemma 1 *If the value* λ_{-1} *is chosen in such a way that the transversality condition,*

$$\lim_{t\to\infty} k_{t+1}\lambda_t \rho^t = 0,$$

holds, then the pair $(\{k_t\}_{t=0}^{\infty}, \{u_t\}_{t=0}^{\infty})$ *which is derived from the dynamical system (3), (6), and (7) is an optimal consumption plan of the North household.*

Let us turn to the representative household in South. For $t = 0,1,2,\ldots$, he solves the following problem[3]

$$\max_{c_{1t}^*, c_{2t}^*} u^*(c_{1t}^*, c_{2t}^*) \text{ subject to } p_t y_1^* \geq p_t c_{1t}^* + c_{2t}^*, \tag{8}$$

where y_1^* is the output of good 1, which is constant, and we assume that u^* is strictly quasi-concave with $2u_1^* u_2^* u_{12}^* - (u_1^*)^2 u_{22}^* - (u_2^*)^2 u_{11}^* > 0$. Notice that the optimal solutions to this maximization problem are given by

$$c_{1t}^* = E_p^*(p_t, v^*(p_t, p_t y_1^*)) \ (\leq y_1^*), \tag{9}$$

$$\text{and } c_{2t}^* = E^*(p_t, v^*(p_t, p_t y_1^*)) - p_t E_p^*(p_t, v^*(p_t, p_t y_1^*)),$$

where v^* denotes the indirect utility function associated with (8).

2.2. The market clearing condition for good 1

With incomplete specialization, we obtain the rental on capital and the wage rate in North as functions of the price of good 1, $r_t = r(p_t)$ and $w_t = w(p_t)$. Let $\kappa_{\min}(p_t)$ ($\kappa_{\max}(p_t)$) denotes the capital/labor ratio in the labor (capital) intensive sector when $r_t = r(p_t)$ and $w_t = w(p_t)$. These factor prices will satisfy full employment for $k_t / l \in [\kappa_{\min}(p_t), \kappa_{\max}(p_t)]$.

Since GDP in North is $r(p_t)k_t + w(p_t)l$ with incomplete specialization, the output of good 1 in North can be expressed as follows:

$$y_{1t} = r'(p_t)k_t + w'(p_t)l. \tag{10}$$

Then, the world market clearing condition for good 1 is

---

[3]In the absence of the international credit market and factor mobility among countries, the South household has to satisfy his budget constraint at each point in time, and hence his problem ends up being a static one.

$$0 = E_p(p_t, u_t) + E_p^*(p_t, v^*(p_t, p_t y_1^*)) - [r'(p_t)k_t + w'(p_t)l + y_1^*].$$

2.3. The dynamic two country model of trade

The dynamical system of our model is

$$k_{t+1} = [1 + r(p_t) - \delta]k_t + w(p_t)l - E(p_t, u_t), \tag{11}$$

$$\lambda_t = \frac{\lambda_{t-1}}{\rho[1 + r(p_t) - \delta]}, \tag{12}$$

$$0 = 1 - \frac{\lambda_{t-1} E_u(p_t, u_t)}{\rho[1 + r(p_t) - \delta]}, \tag{13}$$

$$0 = E_p(p_t, u_t) + E_p^*(p_t, v^*(p_t, p_t y_1^*)) - [r'(p_t)k_t + w'(p_t)l + y_1^*]. \tag{14}$$

For a historically given k_0 and a chosen λ_{-1}, (13) and (14) determine p_0 and u_0. Substituting k_0, λ_{-1}, p_0, and u_0 into (11) and (12), we derive k_1 and λ_0. Repeating a parallel argument, we obtain a time profile $(k_t, \lambda_{t-1}, p_t, u_t)$, $t = 0,1,2,3,\ldots$, which depends on what value is chosen for λ_{-1}.

2.4. Uniqueness of a steady state

The steady state of the dynamical system (11)–(14) is a solution to the system of equations

$$0 = [r(p) - \delta]k + w(p)l - E(p, u), \tag{15}$$

$$1 = \frac{1}{\rho[1 + r(p) - \delta]}, \tag{16}$$

$$0 = 1 - \frac{\lambda E_u(p, u)}{\rho[1 + r(p) - \delta]}, \tag{17}$$

$$0 = E_p(p, u) + E_p^*(p, v^*(p, p y_1^*)) - [r'(p)k + w'(p)l + y_1^*], \tag{18}$$

First, from (16) we see that the steady state rental rate is uniquely determined as

$$r(p) = \frac{1 - \rho}{\rho} + \delta,$$

which is positive and denoted by θ. Then, the steady state price of good 1 is given by

$$\tilde{p} \equiv r^{-1}(\theta).$$

From (15) and (16), we obtain the steady state capital stock as a function of u,

$$\tilde{k}(u) = \frac{\rho}{1 - \rho}[E(\tilde{p}, u) - w(\tilde{p})l]. \tag{19}$$

Substituting (19) into (18) and rearranging it yields

$$\begin{aligned} & y_1^* - E_p^*(\tilde{p}, v^*(\tilde{p}, \tilde{p} y_1^*)) \\ & = -\frac{\rho r'(\tilde{p})}{1 - \rho}\left[E(\tilde{p}, u) - \frac{1 - \rho}{\rho r'(\tilde{p})} E_p(\tilde{p}, u)\right] - \left[w'(\tilde{p}) - \frac{\rho w(\tilde{p})}{1 - \rho} r'(\tilde{p})\right] l, \end{aligned} \tag{20}$$

where the left hand side is the excess supply for good 1 in South, which is constant, nonnegative, and denoted as z_1^*, while the right hand side is the excess demand for good 1 in North, denoted as $-z_1(u)$.

Notice that if both goods are normal, then

$$E_{pu}(\tilde{p},u)>0 \text{ and } \frac{\partial}{\partial u}[E(\tilde{p},u)-\tilde{p}E_p(\tilde{p},u)]=E_u(\tilde{p},u)-\tilde{p}E_{pu}(\tilde{p},u)>0,$$

which implies that

$$E_u(\tilde{p},u)-\frac{1-\rho}{\rho r'(\tilde{p})}E_{pu}(\tilde{p},u)>0 \tag{21}$$

due to the Stolper-Samuelson theorem: $r'(\tilde{p})<0$ holds if good 1 is labor intensive, while $\tilde{p}r'(\tilde{p})/r(\tilde{p})>1$ holds otherwise.

In what follows, we suppose that both goods are normal. Then, z_1 is strictly decreasing (increasing) in u when good 1 is labor (capital) intensive. If $z_1^*=-z_1(u)$ has a solution $\tilde{u}$, which must be unique, the rest of steady state variables are determined from (19) and (17) as

$$k=\tilde{k}(\tilde{u}),$$
$$\lambda=1/E_u(\tilde{p},\tilde{u}).$$

Thus, the steady state with incomplete specialization is unique if it exists.

3. Indeterminacy

3.1. Equilibrium dynamics

Let us linearize the dynamical system (11)–(14) around a steady state to check the number of characteristic roots the absolute values of which are in between 0 and 1. The characteristic equation is

$$f(x)\equiv\begin{vmatrix} 1/\rho-x & 0 & -E_u & z_1 \\ 0 & 1-x & 0 & -\rho\lambda r' \\ 0 & -E_u & -\lambda E_{uu} & -\lambda E_{up}+\rho r' \\ -r' & 0 & E_{pu} & W_p \end{vmatrix}=0, \tag{22}$$

where[4]

$$W_p\equiv\frac{\partial}{\partial p}[E_p(\tilde{p},u)+E_p^*(\tilde{p},v^*(\tilde{p},\tilde{p}y_1^*))-r'(\tilde{p})k-w'(\tilde{p})l-y_1^*]$$
$$=E_{pp}+E_{pp}^*+\frac{E_{pu^*}^*}{E_{u^*}^*}z_1^*-r''k-w''l. \tag{23}$$

One can verify that the first and the second partial derivatives of E have the properties below.

---

[4]Notice that $E_{pu^*}^*z_1^*/E_{u^*}^*$ in (23) can be derived from the following two identities,

$$E_p^*(p,v^*(p,I^*))=-\frac{v_p^*(p,I^*)}{v_{I^*}^*(p,I^*)} \text{ (Roy's identity) and } E^*(p,v^*(p,I^*))=I^*.$$

Lemma 2 *(i)* $E_u = 1/u_2 > 0$; *(ii)* $E_{pp} = -u_2/B < 0$; *(iii)* $E_{pu} = D/Bu_2$; *(iv)* $E_{uu} = H/B(u_2)^3 > 0$, *where* $B \equiv 2\tilde{p}u_{12} - \tilde{p}^2 u_{22} - u_{11}$ *and* $D \equiv u_{12} - \tilde{p}u_{22}$ *are positive due to the strict concavity of* u *and normality in consumption of good 1.*

As a result of straightforward calculations, we get

$$f(x) = \varphi_2 x^2 + \varphi_1 x + \varphi_0,$$

where

$$\varphi_2 = (-W_p H + \Gamma_2 u_2) g(u_2), \tag{24}$$

$$\varphi_1 = -\Gamma_1 u_2 g(u_2) - (\varphi_2 + \varphi_0), \tag{25}$$

$$\varphi_0 = \frac{1}{\rho}\varphi_2 + r' z_1^* H g(u_2), \tag{26}$$

$\Gamma_2 = D(D - \rho r')$, $\Gamma_1 = r'[\rho r' B - (1-\rho)D] > 0$ from (21) and Lemma 2, and $g(u_2) = 1/B(u_2)^2 > 0$. Let x_1 and x_2 be two characteristic roots. As is well known, with $\varphi_2 > 0$ (< 0), the absolute values of x_1 and x_2 are between 0 and 1 if and only if all of $\varphi_2 - \varphi_0$, $f(1)$, and $f(-1)$ are positive (negative).

Notice that it is clear from (26) that there is no possibility of indeterminacy if $z_1^* = 0$: an autarkic free trade equilibrium. This is consistent with the fact that we assume away any market-distortional factor such as factor-generated externalities, imperfect competition, and public goods, which are sources of indeterminacy in a closed economy.

From (24)–(26), we have

$$\varphi_2 - \varphi_0 = \left[\left(\frac{1-\rho}{\rho}W_p - r' z_1^*\right)H - \frac{1-\rho}{\rho}\Gamma_2 u_2\right] g(u_2), \tag{27}$$

$$f(1) = -\Gamma_1 u_2 g(u_2) < 0, \tag{28}$$

$$f(-1) = 2\left[-\left(\frac{1+\rho}{\rho}W_p - r' z_1^*\right)H + \left(\frac{1+\rho}{\rho}\Gamma_2 + \frac{\Gamma_1}{2}\right)u_2\right] g(u_2). \tag{29}$$

Suppose that good 1 is labor intensive ($r' < 0$). Then, we see from (26) that φ_2 must be positive for indeterminacy to arise. However, $f(1)$ must be negative with the normality assumption. Thus, it is necessary for indeterminacy in the case of $r' < 0$ that good 1 is an inferior good, which is equivalent to $D < 0$ from Lemma 2, and $\Gamma_1 = r'[\rho r' B - (1-\rho)D] < 0$.[5]

On the other hand, indeterminacy may occur without inferiority in consumption when good 1 is capital intensive ($r' > 0$). With specific preferences and technologies, in the next section, we will derive a sufficient condition for indeterminacy to arise.

3.2. A specific model

[5]Shimomura (2007) considers the occurrence of indeterminacy when good 1 is labor intensive and inferior.

First, we specify the North utility function as

$$u(c_1,c_2)=\alpha(c_1+c_2)-\frac{\beta}{2}[(c_1)^2+(c_2)^2], \tag{30}$$

where α and $\beta>0$. Then,

$$\begin{aligned}
&u_i=\alpha-\beta c_i>0, \text{ if } c_i<\alpha/\beta \ (i=1,2),\\
&u_{11}=u_{22}=-\beta,\ u_{12}=0,\\
&H=\beta^2,\ B=(1+\tilde{p}^2)\beta, \text{ and } D=\tilde{p}\beta.
\end{aligned}$$

This utility function has properties such that (i) for any positive $\tilde{p}$, the income expansion path converges to the point $(c_1,c_2)=(\bar{c},\bar{c})$, where $\bar{c}\equiv\alpha/\beta$; (ii) H, B, and D are all constant and $u_1=u_2=0$ when $c_1=c_2=\bar{c}$ (or $u=\bar{u}\equiv u(\bar{c},\bar{c})$), both of which make the analysis of indeterminacy much easier. Because with the steady state value of u, $\tilde{u}$, being sufficiently close to $\bar{u}$ (u_2 is sufficiently small at the steady state), we may conclude from (24), (27), and (29) that all of φ_2, $\varphi_2-\varphi_0$, and $f(-1)$ are negative if

$$\frac{1-\rho}{\rho}<\frac{r'z_1^*}{W_p}<\frac{1+\rho}{\rho} \tag{31}$$

holds.[6]

The following Lemma establishes the existence of a steady state with incomplete specialization where $\tilde{u}=\bar{u}$.

Lemma 3 *Let* $\bar{c}>z_1^*$. *Then, there is some value of* l, l_0, *such that* $z_1^*=-z_1(\bar{u})$ *holds and* $\tilde{k}(\bar{u})/l_0\in(\kappa_{\min}(\tilde{p}),\kappa_{\max}(\tilde{p}))$.

On the South utility function, we assume

$$u^*(c_1^*,c_2^*)=\alpha(c_1^*+c_2^*)-\frac{\beta}{2}(c_2^*)^2. \tag{32}$$

Under the quasi-linear utility function (32), we have the excess supply for good 1 in South as follows:

$$z_1^*=\frac{(\tilde{p}-1)\alpha}{\tilde{p}^2\beta} \tag{33}$$

when

$$\tilde{p}>1 \text{ and } y_1^*>\frac{(\tilde{p}-1)\alpha}{\tilde{p}^2\beta}.$$

Finally, technologies in North are assumed to take the following specification.[7]

---

[6]Notice that (31) implies $W_p>0$: the world excess demand for good 1 is increasing in its price at the steady state.

[7]In the case of Leontief technologies, the unit cost functions become linear in w and r as in the following specification. In Shimomura (2007), he assumed $a_1>0$, $b_1=0$, and $a_2=b_2=1$.

$$\chi_i(w,r) = a_i w + b_i r, \quad i = 1,2, \tag{34}$$

where χ_i denotes the unit cost function in sector i and a_i and b_i are constant and nonnegative with

$$\Delta \equiv a_2 b_1 - a_1 b_2 > 0 \text{ and } b_2 < \theta^{-1}. \tag{35}$$

Notice that the first inequality means that good 1 is capital intensive.[8]

From (34) we obtain

$$r(p) = \frac{a_2 p - a_1}{\Delta} \text{ and } w(p) = \frac{b_1 - b_2 p}{\Delta}, \tag{36}$$

which yields

$$\tilde{p} = \frac{a_1 + \Delta\theta}{a_2} \text{ and } w(\tilde{p}) = \frac{1 - b_2\theta}{a_2} \tag{37}$$

due to $r(\tilde{p}) = \theta$.

From (23), (33), (36), and Lemma 2, we have

$$W_p = \frac{(\tilde{p} - 2)\alpha}{\tilde{p}^3 \beta} \tag{38}$$

when $u_2 = 0$, because $u_2^* = \alpha / \tilde{p}$, $B^* = \tilde{p}^2 \beta$, and $D^* = \tilde{p}\beta$, which is derived from (32).

Substituting $r' = a_2 / \Delta$, (33), and (38) into (31) yields

$$\frac{1-\rho}{\rho} < \frac{\tilde{p} a_2}{\Delta}\left(\frac{\tilde{p} - 1}{\tilde{p} - 2}\right) < \frac{1+\rho}{\rho}. \tag{39}$$

This holds if the following conditions are satisfied.

$$\Delta > \frac{a_1 + a_2}{2 - \delta} \tag{40}$$

$$\text{and } \tilde{p} > \frac{2[(2-\delta)\Delta - a_1]}{(2-\delta)\Delta - (a_1 + a_2)} \Leftrightarrow \theta > \hat{\theta} \equiv \frac{2[(2-\delta)\Delta - a_1]a_2}{\Delta[(2-\delta)\Delta - (a_1 + a_2)]} - \frac{a_1}{\Delta}.$$

Based on the above, we have the main result in this paper.

Theorem 1 *Let* $y_1^* > (\tilde{p} - 1)\alpha / \tilde{p}^2 \beta$, $\hat{\theta} < \theta < (b_2)^{-1}$, *and (40) hold. If* l *is smaller than but sufficiently close to* l_0, *then there exists an unique steady state with incomplete specialization, where indeterminacy occurs.*

4. Concluding remarks

We have shown that in a dynamic two country model of trade with technological differences, if a consumable capital is more labor intensive than a pure consumption good, indeterminacy can arise

[8] As will be made clear, the second inequality is necessary to guarantee that the steady state rental rate, θ, is consistent with incomplete specialization.

even under the normality assumption in consumption. This is contrasted with the result in Bond, Iwasa, and Nishimura (2009): Inferiority in consumption is necessary for indeterminacy when there is no technological difference across countries.

5. Reference

Benhabib, J., Q. Meng, and K. Nishimura (2000): Indeterminacy under constant returns to scale in multi-sector economies, *Econometrica* 68, 1541–1548.

Bond, E., K. Iwasa, and K. Nishimura (2009): A dynamic two country Heckscher-Ohlin model with non-homothetic preferences, KIER Discussion Paper Series 686, Kyoto University.

Bond, E. and R. Driskill (2008): Multiplicity and stability of equilibrium in trade models with durable goods, in T. Kamihigashi and L. Zhao eds., *International Trade and Economic Dynamics: Essays in Honor of Koji Shimomura,* Berlin: Springer-Verlag, 281–297.

Nishimura, K. and K. Shimomura (2002): Trade and indeterminacy in a dynamic general equilibrium model, *Journal of Economic Theory* 105, 244–259.

Nishimura, K. and K. Shimomura (2006): Indeterminacy in a two-country dynamic model, *Economic Theory* 29, 307–324.

Shimomura, K. (2007): A Pareto-improving foreign aid in a dynamic North-South model, Chapter 9 in S. Lahiri ed., *Theory and Practice of Foreign Aid,* Amsterdam: Elsevier, 131–153.

04

Adiabatic Quantum Pump

Hisao Hayakawa
Yukawa Institute for Theoretical Physics, Kyoto University

Summary: For open quantum systems we study the adiabatic pump effect, i.e. the existence of the net current under zero average bias during slow modulation of external parameters such as chemical potentials of two reservoirs. Through the analysis based on the quantum master equation, we have confirmed that the appearance of the pump effect is directly related to the existence of geometrical Berry-like phase in the parameter space. This pump effect plays a key role in non-equilibrium thermodynamics, where the entropy production depends on the operating path in the parameter space.

It is well known that the current exists because of the existence of the difference of the voltages or the chemical potentials between two electrodes. However, the electric current can exist even under zero average bias if a system is slowly and periodically modulated by some control parameters such as the chemical potentials. This process is known as the adiabatic pump process. Such an idea on quantum adiabatic pump was firstly proposed by Thouless[1] for a closed system, and was extended to an open system by Brouwer[2]. So far, most of previous studies control system parameters such as the gate voltages, but we need much energy to control such quantities. The objective of this talk is to demonstrate how the pump effect appears if we control the chemical potentials which is easy to control.[3] We also apply such an idea to non-equilibrium thermodynamics, and show that the entropy depends on the operating path.[4] These non-equilibrium currents are associated with the geometric phase analogous to Berry phase.

The method for our analysis is the quantum master equation (QME). For simplicity, we assume that the system for spinless Fermion is weakly coupled with two reservoirs which satisfy the gandcanonical distribution characterized by the temperatures and the chemical potentials. After we tracer over the contribution of the reservoirs under the weak coupling and Born-Markov approximation, we obtain a QME for the reduced density matrix. This QME is generalized as the generalized QME (GQME) by adding a counting field. The average current is obtained from the differentiation of the generating function for GQME, the trace of the reduced density matrix, with respect to the counting field and the zero field limit. The average current can be either the electric current or the entropy production depending on the corresponding choice of GQME. One of the most remarkable results of our analysis is that the average current can survive even under zero average bias, and the existence of the current is directly related to the existence of curvature in the parameter space of the control parameters such as the chemical potentials.

Our findings on the adiabatic pump of the electric current for spinless Fermions are as follows. (i) There is no net electric current for a system without interactions, and (ii) the pump effect appears as a result of the repulsive interaction between electrons, which is characterized by U. The main results are summarized in Fig.1.

The calculation for the adiabatic pump is directly applicable to the calculation for the entropy production in non-equilibrium system. So far, we have only published the result for a classical Markovian jump process[4], but we also have a result for QME. Similar to the case of the adiabatic pump for the electric current, we found that the entropy production depends on the operating path in the parameter space. This entropy production is reduced to the conventional generalized Clausius

entropy.

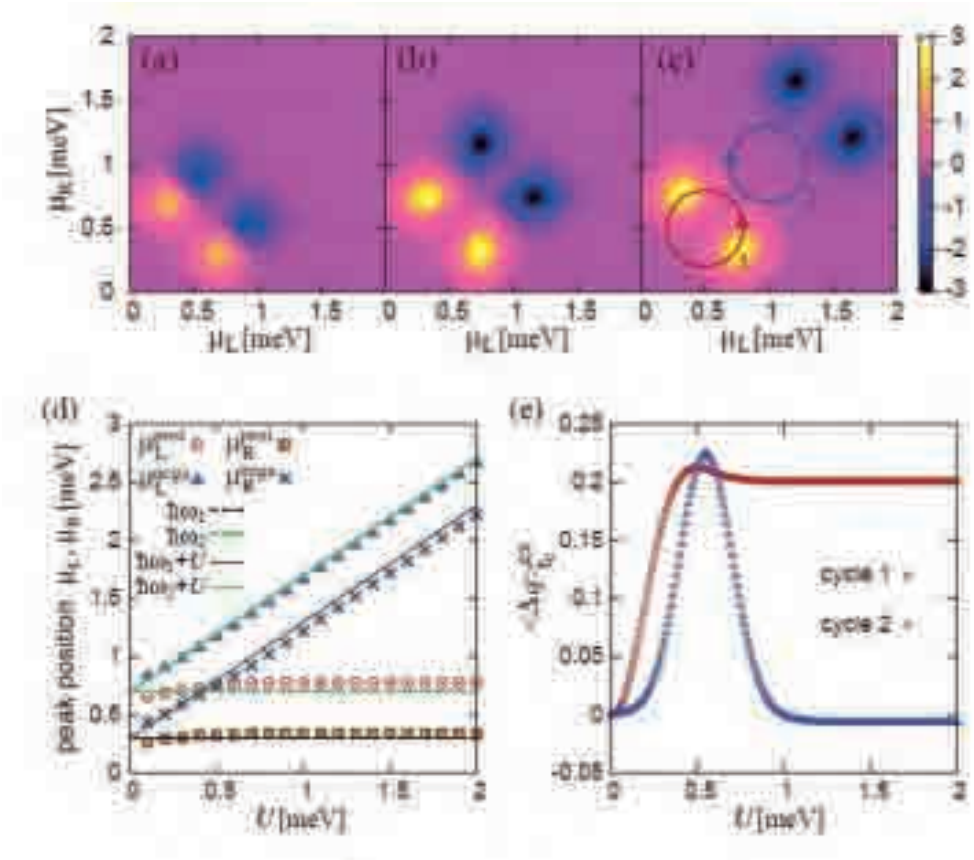


Fig.1 (a)-(c) The Berry-phase like curvature for the interacting spinless electrons with U=0.25 meV (a), U=0.5 meV (b) and U=1.0meV (c). (d) The peak positions of the curvature in the paramerer space, where the negative peaks are proportional to U. (d) The net electric current for the cycles 1 and 2 in Fig. 1 (c).

We also have preliminary results for the adiabatic quantum pump for Kondo system, in which the spin of electrons play essential roles. Moreover, we have also obtained some preliminary results for non-adiabatic pump for a spin-boson system, in which the Berry-like phase is extended to the dynamical phase, and the result strongly depends on the initial condition.

In this talk, I would like to present our results mentioned above as well as the stress of possible extension to the novel concept in non-equilibrium thermodynamics.

Acknowledgment: This talk is based on the collaboration with T. Sagawa, T. Yuge, A. Sugita, R. Yoshii, K. Watanabe and C. Uchiyama. I would like to express my sincere gratitude to my collaborators.

Reference

[1] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).
[2] P. W. Brouwer, Phys. Rev. B **58**, R10135 (1998).
[3] T. Yuge, T. Sagawa, A. Sugita and H. Hayakawa, arXiv:1208.3926.
[4] T. Sagawa and H. Hayakawa, Phys. Rev. E **84**, 051110 (2011).

05

The emergence of the chemical bond

Stewart Clark

Department of Physics, University of Durham, Science Labs, South Road, Durham DH1 3LE, UK

1 Introduction

One of the most challenging problems in condensed matter physics today is the description of the properties of materials from first principles, that is, without the need for experimental information. The immediate difficulty arises from the need to describe the structure of the material in a means that gives in sight into the properties of the material from a simple, conceptual viewpoint. Can we take a take a simple description of a material and give details of it's properties?

The chemical bond is one such method of describing a material. With a basic background in chemistry one find terms such as covalent bond, ionic bond and metallic bonding. A little further and we come across terms such as hydrogen bond, molecular bonding and van der Waals bonds. These terms alone gives us some insight into some of the properties of materials -

- Covalent bond: high melting point, good thermal insulator, bad electrical conductor, brittle
- Metallic bond: high melting point, good electrical and thermal conductor, ductile
- Ionic bond: High melting point, electrical insulator while solid but conductor in melt

and so on. However there are (at least) two problems with this. Firstly these concepts are purely qualitative and secondly the definitions of such bonds are actually very badly defined; there are many (in fact, most) material's bonding which are either intermediate between various pairs of these definitions or have a mixture of these bonds. Additionally, one usually infers the type of bonding in a material from its properties rather than the other way round.

The question arises, given only fundamental principles in Physics, can we do better? Can we determine the properties of materials? Can we find a model in which the chemical bond is a well-defined, emergent property rather than a conceptual one imposed after properties are known?

Yes. Quantum mechanics is used, but it is only in recent years with the onset of fast computers (see Figure 1), advanced theory and algorithmic methods that it has become possible. We use first principles computer simulations.

2 Modeling and quantum mechanics

The simulator builds a model of a real system and explores its behaviour. The model is a mathematical one and the exploration is done on a computer, and in many ways simulation studies share the same mentality as experimental ones. However, in a simulation there is absolute control and access to detail, the ability to compute almost any observable, and given enough computer muscle, exact answers for the model. These strengths have been exploited for the last sixty years and have led to many advances in the theory of condensed matter. However, it is only in the last twenty years or so that we have been able to compute the properties of condensed matter from first principles[1].



Figure 1: Accurate calculations to obtain quantitative data on emergent properties of quantum mechanics need large compute power such as the UK national supercomputing facility shown here. Some calculations in this paper are performed on it.

The first-principles approach is vastly ambitious because its goal is to model real systems using no approximations whatsoever. That one can even hope to do this is down to the accuracy of quantum mechanics in describing the chemical bond. Diracs apocryphal quip that after the discovery of quantum mechanics the rest is chemistry sums it up: if one can solve the Schrödinger equation for somethingan atom, a molecule, assemblies of atoms in solids or liquidsone can predict every physical property. Perhaps Dirac didnt realise quite how difficult doing the rest is, and it has taken great effort and ingenuity to take us to the point of calculating some of the properties of materials with reasonable accuracy[2]. However, to focus limitations is to miss the point: the impact of simulations on our thinking about condensed matter problems is immense.

Quantum mechanics provides a reliable way to calculate what electrons and atomic nuclei do in any situation. The behaviour of electrons in particular governs most of the properties of materials. This is true for a single atom or for assemblies of atoms in condensed matter, because quantum mechanics describes and explains, and allows us to define, chemical bonds. Therefore we can understand the properties of any material from first-principles, that is, based on fundamental physical laws and without using free parameters, by solving the Schrödinger equation for the electrons in that material. This, however, is a tall order. We rapidly run into difficulty because electrons interact strongly with each other. The alarming consequence is that exact pencil-and-paper solutions exist only for a single electron in exceptionally simple situations: solving the Schrödinger equation for the hydrogen atom is a classic undergraduate task, but solving it for helium requires advanced theoretical and computational approaches. The problem of interacting electrons in condensed-matter physics, one manifestation of the many-body problem, is the defining challenge of the subject.

Mathematically, the problem to be solved is

$$\left[\frac{1}{2}\sum_{I=1}^{N}\frac{1}{M_I}\nabla_I^2+\frac{1}{2}\sum_{I=1}^{N}\frac{1}{m}\nabla_I^2+\frac{1}{2}\sum_{I\neq J}^{N,N}\frac{Z_I Z_J e^2}{|R_I-R_J|}+\frac{1}{2}\sum_{i\neq j}^{n,n}\frac{e^2}{|r_i-r_j|}-\sum_{I}^{N}\sum_{i}^{n}\frac{Z_i e^2}{|R_I-r_i|}\right]\Psi\left(\{R_I\},\{r_i\}\right)=E\Psi\left(\{R_I\},\{r_i\}\right)$$

where capital letters denote atomic nuclei (R is position, M mass and Z atomic number) and lower case letters denote electrons. Here, the symbol Ψ is the famous quantum mechanical wavefunction in which all information about a system is stored. I will not go into any details

here, other than to simply state that (i) even though this equation has been known for many years the mathematical and computational tools to solve it have only been known recently and (ii) from this equation all the properties of matter (materials) emerge.

Density-functional theory[3, 4] takes a remarkable approach to solving this complex mathematical and computational problem. It is both a profound, exact theory for interacting electrons, and a practical prescription calculating a solution. Its contribution in both these respects received the highest recognition with the award of the Nobel prize for chemistry in 1998 to Walter Kohn. The beauty of DFT is that one makes no attempt to compute the many-body wavefunction, but instead aims to calculate the electron density (the probability of finding an electron at a particular point in space). This puts the electron change density on a firm physical footing from a quantum mechanical point of view, and remarkably overriding the need for the wavefunction. From this we are able to define and calculate the emergent properties, and in particular, the chemical bond.

3 Algorithmic and computational tools

The author of the current paper is a co-author of the CASTEP programme[1, 2], which is a first principles quantum mechanical code for performing electronic structure calculations, as described above. Within the density functional formalism it can be used to simulate a wide range of materials including crystalline solids, surfaces, molecules, liquids and amorphous materials; the properties of any material that can be thought of as an assembly of nuclei and electrons can be calculated with the only limitation being the finite speed and memory of the computers being used. These are all emergent properties of the defining equation, above. Rather than detailing the methods further we now present some specific examples.

4 Examples of emergent properties

4.1 The 'Classical' Covalent Bond

After many years of study, the polymorphism of group IV elements remained unresolved. This is due to the existence of metastable phases (phases which are stable, but high in energy). These elements exist naturally in the diamond structure; we show this along with the electron charge density in Figure 2. In this Figure, it is immediately apparent that the covalent bond has emerged. The interplay of attraction and repulsion between electrons and the atomic nuclei along with their kinetic energy essentially produce this complex result from nothing but these fundamental interactions.

However is we take a crystal of diamond-silicon and put it under modest pressure the result is remarkably different. It turns this prototypical semiconductor into a metal. The electronic structure completely changes and hence the material properties are completely different. It still contains the same number of silicon atoms, the same number of electrons, but the effect of pressure is to change the position of the atomic nuclei and silicon is no longer the same material. It is metallic and reflective, like any normal metal.

On depressurisation from the (first!) high pressure metallic phase, it does not return to the initial diamond phase, but form complex 4-fold coordinated high density structures (known as BC8 and ST12)[5, 6]. But again we find it to be covalently bonded and shown in Figure 3. Of course, because this is done from first principles we can be quantitative about the results. We

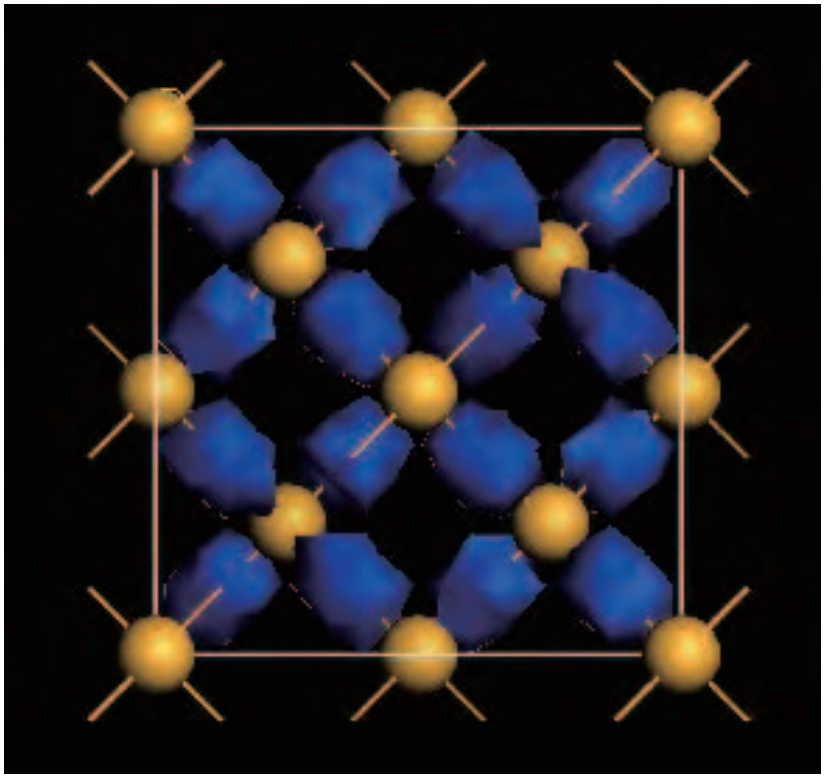


Figure 2: Diamond structure and electronic charge density of silicon. The orange spheres represent the positions of Si atoms, while the blue objects show the volumes of high electron charge density. These emergent structures are covalent bonds.

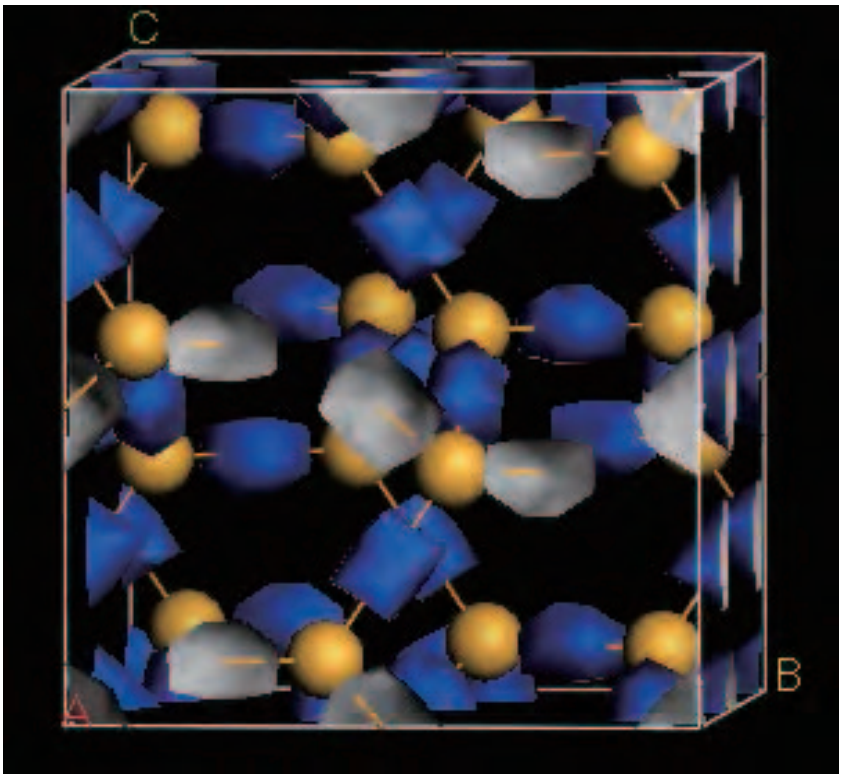


Figure 3: Silicon in the covalently bonded BC8 structure.

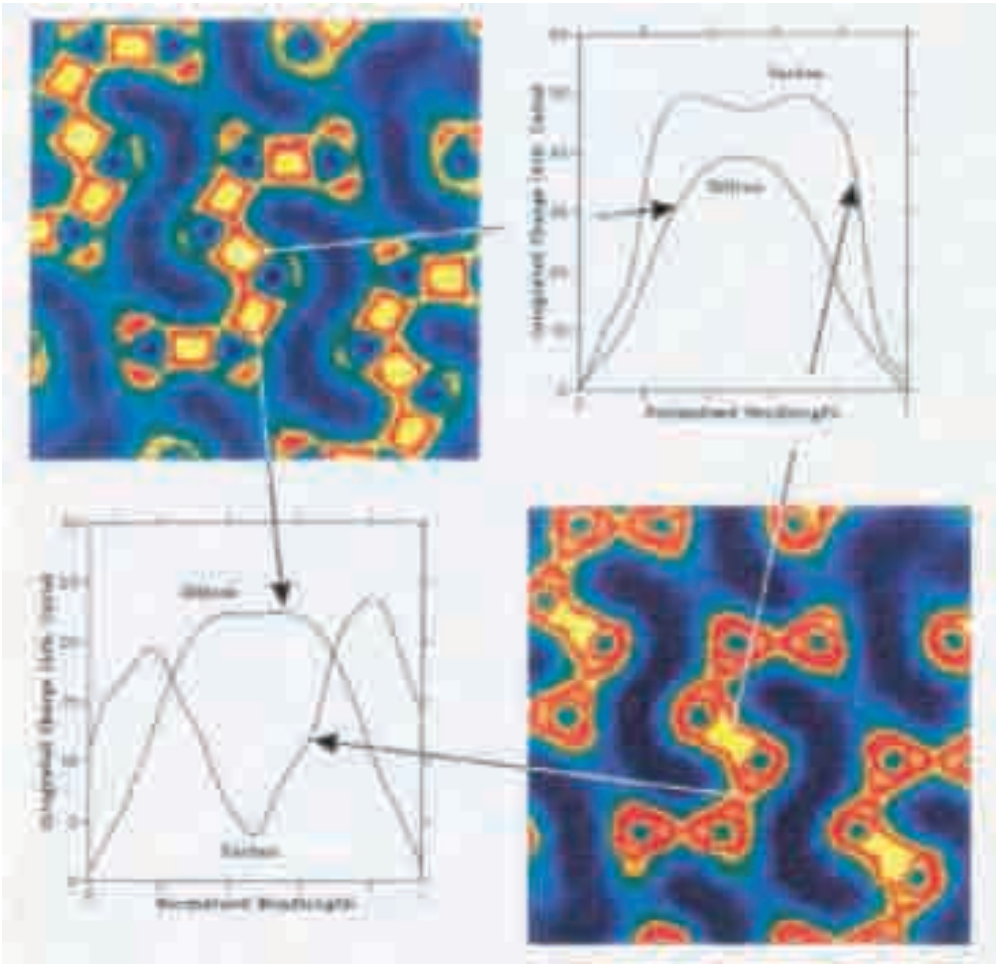


Figure 4: Analysis of the emergent covalent bond in carbon and silicon in different structures. The structures within a covalent bond can be seem and quantified. The properties of the covalent bond is complex and differs from material to material, structure to structure.

can examine the binds in detail, at all points in space. We can even compare from material to material, element to element. Examples of such quantitative results is given in Figure 4.

The structures of silicon and carbon contain a rich mixture of possible bonding topologies; under some conditions covalent, under other conditions metallic.

4.2 Defects in ZnO

The importance of semiconductor point-defects in controlling the properties of semiconductors and hence the technological devices that contain them has long motivated a strong interest in predicting their properties by quantum calculations[7]. The emergent properties from the Schrödinger equation have proved massively influential in the design and manufacture of semiconducting devices and have undoubtedly played a role in the design of the devices that you are now familiar with[8].

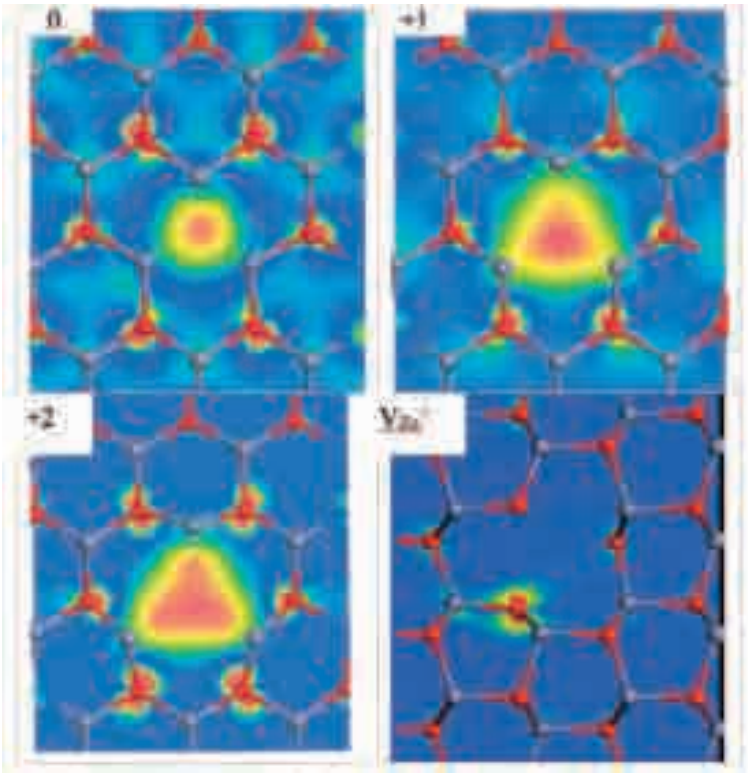


Figure 5: The unexpected 3-centre orbital (a new type of covalent bond) that emerged. They have also been seen in a range of materials such as carbon and boron.

The three main quantities of interest are (i) The defect formation energy which sets the defect concentration that can be attained in the solid at a given temperature in equilibrium, (ii) the energy required to ionize (add or remove charge) the centre from the defect state and (iii) the spatial localization of the defect states. These are important in device technologies and they all rely in a quantitative understanding of the chemical bond. Investigations of these show a massive range of bonding topologies and types as well as a few new surprises which emerged, unexpectedly from the Schrödinger equation.

In Figure 5 we show the bonding configuration of the defect state responsible for the technological properties of ZnO[9]. The standard type of bond in metal(!)-oxides are a combination of ionic and covalent (not metallic) with electrons partially shared and partially donated between pairs of atoms. However, for the defect state, the bond is found to be a covalent-like bond but, unusually instead of being a increase in charge between a pair of atoms in the traditional way, it is found to be a three-centre orbital. Such bonding configurations do not fall into the standard categories of chemical bonding and is an unusual emergent property of the Schrödinger equation. Other interesting bonds are when fractional amounts of electrons are shared between atoms rather than the expected integer numbers.

5 Summary

It has been possible to model such a wide range of emergent electronic structures because of advancement in first principles, theoretic and algorithmic methods as well as the significant compute power available today. It would not be possible to find most of the electron configurations by empirical means and hence qualitatively investigate the emergent properties of the Schrödinger equation, namely the concept of the chemical bond.

References

[1] Segall, MD; Lindan, PJD; Probert, MJ; Pickard, CJ; Hasnip, PJ; Clark, SJ; Payne, MC, " First-principles simulation: ideas, illustrations and the CASTEP code", JOURNAL OF PHYSICS-CONDENSED MATTER, Vol 14, Issue 11, Page 2717-2744 (2002).

[2] Clark, SJ; Segall, MD; Pickard, CJ; Hasnip, PJ; Probert, MJ; Refson, K; Payne, MC, " First principles methods using CASTEP", ZEITSCHRIFT FUR KRISTALLOGRAPHIE, Vol 220, Issue 5-6, Page 567-570 (2005).

[3] Hohenberg, P; Kohn, W, PHYSICAL REVIEW, Volume 136, Page B864 (1964).

[4] Kohn, W; Sham, LJ, PHYSICAL REVIEW, Vol 140, Page A1133 (1965).

[5] Crain, J; Clark, SJ; Ackland, GJ; Payne, MC; Milman, V; Hatton, PD; Reid, BJ, " Theoretical study of high-density phases of covalent semiconductors .1. ab-initio treatment", PHYSICAL REVIEW B, Vol 49, Issue 8, Page 5329-5340 (1994).

[6] Clark, SJ; Ackland, GJ; Crain, J, " Theoretical study of high-density phases of covalent semiconductors .2. empirical-treatment", PHYSICAL REVIEW B, Vol 49, Issue 8, Page 5341-5352 (1994).

[7] Liu, D; Clark, SJ; Robertson, J, " Oxygen vacancy levels and electron transport in Al2O3", APPLIED PHYSICS LETTERS, Vol 96, Issue 3, Article No. 32905 (2010).

[8] Robertson, J; Clark, SJ, " Limits to doping in oxides", PHYSICAL REVIEW B, Vol 83, Issue 7, Article No. 75205 (2011).

[9] Clark, SJ; Robertson, J; Lany, S; Zunger, A, " Intrinsic defects in ZnO calculated by screened exchange and hybrid density functionals", PHYSICAL REVIEW B, Vol 81, Issue 11, Article No. 115311 (2010).

06

Toward Unveiling the Explosion Mechanism of Gamma-Ray Bursts

Shigehiro Nagataki
Yukawa Institute for Theoretical Physics

Summary: Gamma-Ray Bursts are the most energetic explosion in the universe. Their explosion mechanism is not known well. We are trying to understand the extreme phenomena using the state-of-the-art physics. We also adopt the approach of numerical simulations using super-computers to unveil the explosion mechanism.

Introduction:

Gamma-Ray Bursts (GRBs) are the most energetic explosion in the universe. If you see the universe in gamma-ray band, you will find a bright band of Milky Way. This is because the earth is in Milky Way. However, once a GRB happens in the universe, it becomes the brightest object even though it is very far away from us. This is because GRBs are very energetic and bright objects. The duration of GRBs is typically 10 sec or so. The explosion energy of GRBs is 10 times greater than the total energy that the sun emits during its whole life.

The first report of the observations of GRBs was done in 1973. After that, it had been debated what was the origin of GRBs. About 25 years later, in 1997, X-rays are detected after a GRB, which made it easy to localize the position of the GRB. Follow up observation in optical band for the GRB was done, and a host galaxy of the GRB was identified. One year later, on Apr. 25th 1998, a GRB that was associated with a supernova explosion was found. Since then, such association has been confirmed by many examples. Nowadays it is believed that (at least, a part of) GRBs are born from supernova explosions.

It has turned out that the supernovae associated with GRBs are also very energetic. The typical explosion energy of the supernovae is more than 10 times the explosion energy of a typical supernova. Thus it is believed that something special is happening at the center of the progenitor stars that produce GRBs associated with energetic supernovae (sometimes it is called as hypernovae). However, it is still unknown what is happening at the center.

What is something special? One possibility is that a black hole is formed at the center as a result of gravitational collapse. During the life of massive stars, gravitational force is supported by thermal pressure which is produced by nuclear reactions. However, at the end of their lives, flesh nuclei have been run out, and enough thermal pressure cannot be produced by nuclear reactions. As a result, gravitational collapse happens. Then a black hole may be produced at the center. This is a promising scenario called as a collapsar scenario. Especially, when the progenitor star is rapidly rotating, the resulting black hole will be also spinning very fast. In this case, if the black hole is surrounded by proper electro-magnetic fields, the rotation energy of the black hole can be extracted. This is pointed out by Bandford and Znajek (1977). This Blandford-Znajek (BZ) mechanism may be working efficiently in a collapsar.

The BZ mechanism is promising, and I am interested in it. However, in the original paper by

Blandford and Znajek (1977), they assumed a monopole like magnetic fields. Also, their formulation is applicable only for a slowly spinning black hole. In the case of GRBs, however, it is desirable to consider a rapidly spinning black hole since a lot of rotation energy has to be extracted from the rapidly spinning black hole to explain the huge amount of energy of GRBs. Also, the assumption of monopole like configuration of magnetic fields seems to be unrealistic. Thus we need to search for another approach to investigate the BZ mechanism in a collapsar. That is why I decided to develop a General Relativistic Magneto-Hydro Dynamic (GRMHD) code. By developing it, we can investigate the effect for rapidly spinning black hole surrounded by any configuration of electro-magnetic fields. I have developed the GRMHD code from scratch, then I named it as YukAwa institute's MAgneTO-hydro (YAMATO) code (YAMATO means the old name of Japan).

Using the YAMATO code, I have performed GRMHD simulations of collapsars. There are two groups including me that have reported such simulations. Both of us have done 2-dimensional GRMHD simulations assuming axisymmetry (Barkov and Komissarov 2008; Nagataki 2009). Due to the BZ mechanism, energetic jet is launched along the spin axis of the black hole in both groups. It is important that a jet is launched as a result of their simulations, because it is suggested theoretically and observationally that GRBs are jet-like phenomena. Recently, I have performed 3-dimensional GRMHD simulations (Nagataki 2012). I have found that the result of 3-dimensional simulation is almost same with the 2-dimensional one if the initial condition is axisymmetric. On the other hand, however, the system becomes unstable if small perturbation is introduced in azimuthal direction. In this case, the result becomes completely different from the 2-dimensional simulation. Especially, it has been found an expanding outflow is propagating outwardly. This component may be responsible for a supernova component while the jet is responsible for a GRB. Further analysis and simulations are necessary to confirm whether this picture is right or not.

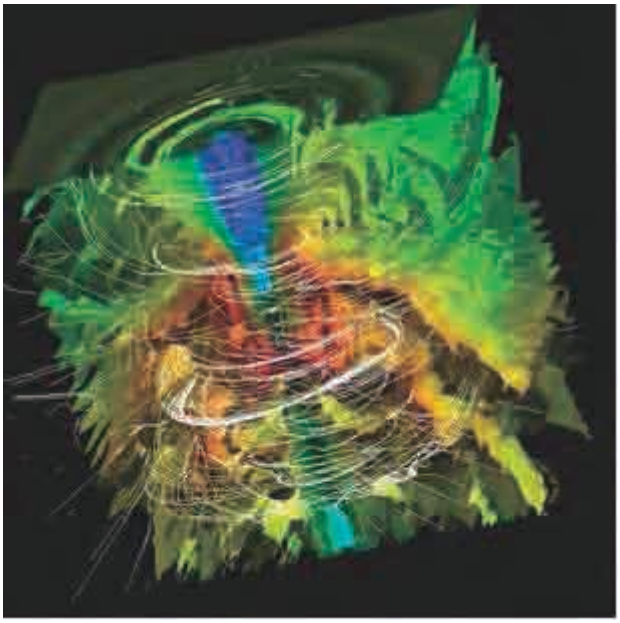


Fig.1 Example of 3-dimensional simulation with axisymmetric initial condition. Jet is launched by the BZ mechanism.From this simulation, as long as the initial condition is axisymmetric, it is concluded that dimensionality does not change the story: (i) steady radial inflow phase, (ii) amplification of magnetic fields due to the differential rotation, (iii) the BZ effect starts to work and a jet is launched along the rotation axis.

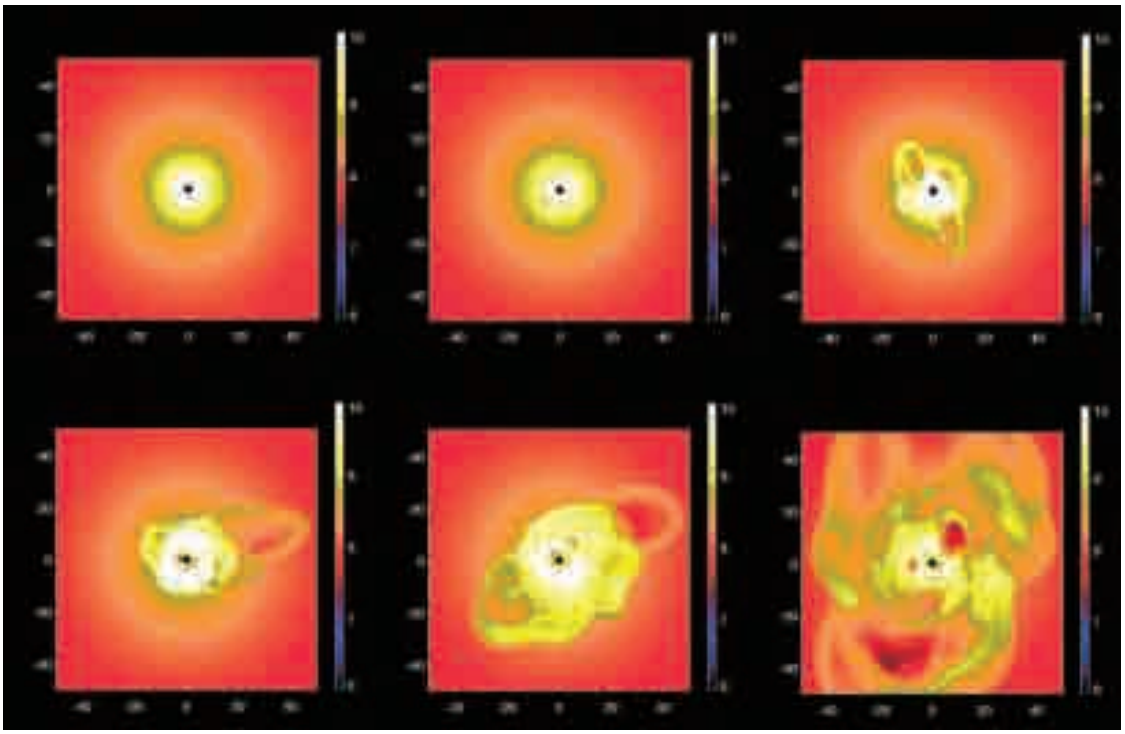


Fig.2 Slices of 3-dimensional simulations with perturbation in density in azimuthal direction. The dynamics evolves with time from top-left to bottom-right. Color contours represent rest mass density in logarithmic scale. It is confirmed that the system is unstable for the density fluctuation in the azimuthal direction and the resulting dynamics is different from axisymmetric simulations. The driven outflow may be related with the supernova component.

Examples of 3-dimensional simulations are shown in Fig.1 and 2. Explanation on the simulation conditions are written in the figure captions, respectively. The jet seen in Fig.1 will be related with a GRB, while the expanding outflow seen in Fig.2 may be related with a hypernova component. 3-dimensional simulations on collapsars have been done for the first time, and further investigation is necessary to unveil the explosion mechanism of GRBs. In my talk, I would like to present the current understanding on the mechanism and future prospect for the study on that.

Reference :

Barkov and Komissarov 2008, MNRAS, 385, L28
Blandford and Znajek 1977, MNRAS, 179, 433
Nagataki 2009, ApJ, 704, 937
Nagataki 2012, ApJ, submitted.

07

Quantum physics from emergence: materials, experiments and calculations

Tom Lancaster
Department of Physics, Durham University

The concept of emergence allows the physicist to explore materials in the laboratory whose properties exactly mirror phenomena from across science. In this way the interactions of fundamental particles, the nature of the early universe and the mathematical description of quantum mechanical fields may all be explored in real materials in the laboratory. Here I describe some examples of the emergent properties of exotic matter and our attempts to understand them with experiment and theory.

The discussion of emergence in quantum mechanics (and in particular in condensed matter physics, which describes the quantum physics of solid and liquid matter) relies on two central concepts: **broken symmetry** and the **renormalization group**. Here I will provide a brief introduction to these ideas and list some examples of their consequences for the measurements we make in the laboratory.

Broken Symmetry

Particularly important to the emergentist view of the physical world is the concept of *broken symmetry,* which explains the existence of distinct phases of matter, the transitions between these phases and the black hole-like barriers that separate them. The notion of symmetry breaking is central to modern condensed matter physics [1,2]. The archetypal example of this phenomenon is found in the thermodynamics of a magnet [3]. A magnet is a piece of matter with an arrow, which we will call a “magnetic moment” placed at regular intervals in a three-dimensional lattice. At high temperature these magnetic moments are disordered, which is to say that they point in random directions. This state is known as a *paramagnet*. It's perhaps obvious, but rather important to notice that the physics of the system is the same if we rotate all of these moments through some angle. They pointed in random directions before the rotation and they point in (different) random directions after the rotation. However, a device that measures the effect of the magnetic moments (a “magnetometer”) would not be able to tell that we have rotated the moments. Their effects cancelled out before the rotation (since they're pointing in all possible directions) and will cancel out afterwards. This invariance of the properties of the system to rotations is known as a **symmetry**.

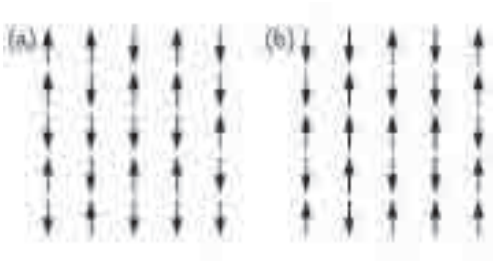


Fig. 1: (a) The paramagnet. (b) The same paramagnet with all moments rotated. Note that we cannot tell the difference between (a) and (b) in a typical experiment.

Upon cooling through a critical temperature T_c, the magnetic moments spontaneously line up along a unique direction (for iron T_c is 770 °C). This state is known as a *ferromagnet* (and the change between paramagnet and ferromagnet is known as a *phase transition*). The properties of the ferromagnetic system measured by the magnetometer will now change if we rotate all of the moments through the same angle. Before the rotation all moments pointed along one direction,

now they all point along another. We say that the symmetry has been spontaneously lowered, or "broken", on cooling through T_c. An important technical point arises when one makes even the simplest a mathematical model of this physics. The phases on either side of a symmetry breaking transition are separated by a *mathematical singularity* in the energy of the system. A singularity makes it impossible to continuously track the properties of the system through the transition and infer the properties of the broken symmetry state from the more symmetrical state. This is the feature that leads to the description of the properties of the broken symmetry state in terms of emergence. Interestingly, there is a close analogous between the mathematical singularity in the phase transition and the black hole in cosmology where a black hole singularity prevents observers gaining knowledge of what's inside the hole. This has led some researchers to call phase transitions black holes in materials!

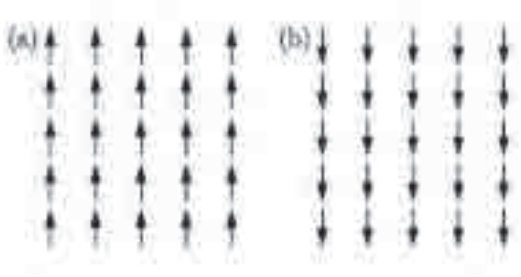


Fig. 2: (a) the ferromagnet with all moments lined up. (b) The ferromagnet with all moments rotated. Notes that (a) and (b) may be distinguished in an experiment.

The presence of a mathematical singularity in the description of a phase transition creates a fundamental barrier to our knowledge of the broken symmetry state. The properties of the system after symmetry breaking cannot be predicted from the high-temperature state and it is in this sense that these properties are *emergent*. This is a significant statement and implies that even if it were possible to exactly solve the Schrödinger equation for every atom in a vessel full of steam (an impossible problem with our current computational capabilities!) the solution would be of limited use since there is nothing to guarantee that the solution would describe the properties of the ice that would form if we cooled the vessel down in a freezer.

The description of the phase transition in a magnet described above was first formulated by the great 20th century physicist Lev Landau and later popularized by Nobel laureate Philip Anderson. It's popularity stems from the fact that it may also be used to classify an enormous number of physical phenomena (including the distribution of Earthquakes, the formation of snowflakes, the length of polymer chains and even the nature of the early Universe) and, as we shall see, reveals a class of emergent phenomena. Anderson notes that the breaking of symmetry leads to four properties of a system, which we examine below.

- **Phase transitions**

Cooling through the critical temperature causes a *phase transition*. Like the familiar change of water into ice, the paramagnet to ferromagnet phase transition has a number of measurable consequences. Perhaps the most dramatic feature of phase transitions is the concept of *universality*: phase transitions have a number of basic similarities that are independent of the microscopic detail of what's going on. As a consequence there are a number of basic similarities between the transitions involving liquids and solids, paramagnets and ferromagnets, metals and superconductors and still more exotic examples. This is the sense in which broken symmetry unifies an enormous amount of physics.

- **Excitations**

If we give a very cold solid a little energy something remarkable happens: we create quantum mechanical entities in the materials that act like microscopic particles. (Note that these particles should not be identified with the atoms forming the solid, they are quite different.) We say that the system is in an excited state and there is a spectrum of particle-like excitations. The first truly emergent phenomenon we encounter is found when we give a little energy to the system in its broken symmetry state. We find that upon symmetry breaking new quantum mechanical particle excitations are present that were not there in the unbroken symmetry state. For the magnet these new particles are called *magnons*.

- **Rigidity**

The next emergent property we find is known as rigidity and is an extension of the well-known rigidity of solids, which allows one to pick up a ruler by only holding one end. Once a symmetry is broken, attempts to change the underlying configuration lead to new forces appearing. This is *generalized rigidity* and cannot be predicted from the underlying equations describing the constituent parts of a solid. The new, rigid forces lead to a magnet showing permanent magnetism: a magnet will attempt to keep all of its moments pointing along its chosen direction and forces will appear to maintain this configuration if an attempt is made to change it.

- **Defects**

In an extended system it is possible to break symmetry in different ways in different parts of the material. For example all of the moments on the left hand side of a material might point up and those on the right might point down. The boundary between these regions is known as a defect. For the magnet the defect is known as a domain wall. A defect has a basic mathematical similarity with a quantum mechanical particle and these defects may be thought of as extending the types of excitation that may exist in a solid. Again these are emergent entities, which owe their existence to the breaking of symmetry.

The picture of broken symmetry described here relies on things happening as we change temperatures. Such models are, roughly speaking, called classical descriptions (as opposed to quantum mechanical). The classical picture may be extended to a quantum mechanical one (that takes place at zero temperature) where symmetry may be broken by tuning pressure or some other external stimulus, rather than temperature. This leads to the notion of a “quantum phase transition” which has been the focus of recent research effort.

Renormalization group

Despite the limitations caused by mathematical singularities, our understanding of the physics of broken symmetry was revolutionized in the 1970s by the development of the *renormalization group* (RG) [1] by Kenneth Wilson (and many others). The RG is a set of mathematical tools that allow detailed, accurate calculation of some of the properties involved in spontaneous symmetry breaking by examining how physics is described on different *length scales*.

The concept of a particular length scale becomes rather vivid when we think about how we make measurements on a system. We consider a gas of uranium atoms. If we are interested in the sound waves transmitted by the gas then our length scale of interest is typically centimeters. (The wavelength of audible sounds is between 20 m and 20 mm.) We don't care about details on the nanometer scale (where we see the structure of the atom) and neither do we care about details occurring on the scale of kilometers. In contrast, if we are interested in the electron motion in the atom our length scale of interest is now nanometers (i.e. 10^{-9} m). If we become interested in instabilities in the nucleus of the uranium then our length scale of interest becomes the

femtometer (10^{-15} m). Clearly, if we are interested in one length scale then details of the physics that occur on smaller (or larger) scales are not necessarily of interest to us. However we will presumably feel their average effect in some way. The RG represents the mathematical method to track the physics of a system as we vary the length scale to the one of particular interest to us. As might be expected, it involved a certain amount of averaging over the small-scale properties of a system.

In practice, the RG method involves taking a complex, microscopic description of a system and identifying a set of parameter that describe how the elements of the system couple together. These are known as *coupling constants* and might include the electric charge of a particle or its mass. The key here is that although we are trained to think of these parameters as constants of nature, in fact they are not! These "constants" actually take on different values depending on the length scale at which we're working. (An example of this strange state of affairs is considered below.) This is the key to understanding the RG. We then mathematically average out detail that occurs over the shortest length scales and chart the evolution of our couplings. This process is repeated, removing more and more detail, and results in the couplings approaching limits which have been found to agree remarkably well with those derived from experimental measurements that take place at the length scale of interest. In fact, the use of the RG extends well beyond symmetry breaking and is central to our understanding of a remarkable array of emergent phenomena including localized states in metals (where electrons get stuck in a particular position), *topological* transitions involving vortices in magnets and the physics of asymptotic freedom which governs the behavior of subatomic quarks. We describe a selection of these in the next section.

Examples

To illustrate the central concepts of broken symmetry and the renormalization group we consider some examples from Nature that may be studied experimentally by a condensed matter physicist.

- **The solid**

The most familiar example of a broken symmetry state is a solid. At high temperatures atoms form a liquid. As the temperature is lowered the atoms freeze into a regular repeating pattern we recognize as a crystalline solid. The emergent quantum particles of a solid are known as phonons. These are packets (or quanta, if you prefer) of vibrational energy that have the properties of particles. The rigidity of a solid is familiar to anyone who has stubbed their toe! An attempt to deform the solid by moving its atoms away from their perfect pattern is met with the occurrence of a new force (possibly acting on the foot). The defects in a solid may be seen in high-resolution photographs of crystalline materials. In these it is possible to see mismatches in the arrangement of atoms showing how symmetry has been broken in different ways in different parts of the solid.

- **Superfluidity and Superconductivity**

The superfluid state involves the dissipationless flow of particles and is found at very low temperatures in liquid helium. That is to say atoms of helium flow without any friction. If one tries to pick up such a liquid in a bucket it drains out over the sides immediately! Superconductivity is similar in many respects to superfluidity and involves the dissipationless flow of charged particles at very low temperatures (typically a few degrees above absolute zero for many metals). That is to say, charge flows in a superconductor without electrical resistance and there is hope that materials showing such superconducting states may solve the energy crisis. Superconductivity and superfluidity look a lot like perpetual motion realized in the laboratory! (Honesty compels me to tell you that this almost, but not quite, the case.) The symmetry that is broken in the case of the superfluid and superconductor is more subtle than the ones we have met

so far, but arguably is more fascinating! In the spirit of the examples above we might ask what exactly is lining up. The answer is a *quantum mechanical wavefunction*.

Superfluidity

Let's consider a superfluid. The quantum wavefunction is a mathematical object that describes all of the known physical content of the superfluid. It may be thought of as a machine that inputs a position and outputs two numbers: the first tells us how many particles are present at that position and the second tells us a number called the *phase* of a wavefunction. (Perhaps surprisingly, this is all that is needed to completely characterize a superfluid!) As with magnetic moment in the example of a magnet, the phase may also be visualized as an arrow. The wavefunction therefore tells us that at a particular point where we expect the find a particle, the arrow points in some particular direction.

At temperatures above the superfluid transition we might ask about the phase at a number of positions and we find that it takes random values (just like the paramagnet). If we cool the superfluid through the superfluid transition we find that the arrows all align (just like the ferromagnet). A consequence of this (which is certainly not obvious from the preceding discussion) turns out to be that the particles making up the superfluid lose their identities and we are no longer able to say exactly how many we have at a particular point.

If we now ask how the superfluid fits into the scheme above. We note that the emergent particles are called Bogolons and represent waves of indistinct atoms emerging from the superfluid. The emergent rigidity phenomenon is the most dramatic property of the system: the flow of particles without resistance. An attempt to change the phase at one end of the sample results in a superflow of particles, which acts to cancel out this anomaly in the phase. The emergent defect of a superfluid is the vortex. This is a pattern of phase that winds round in a circle as shown below.

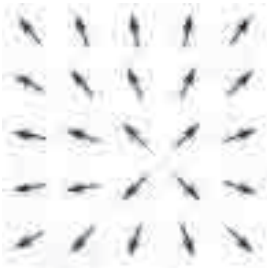


Fig. 3: The vortex defect in a superconductor.

Superconductivity

The superconductor is similar to the superfluid except that the particles of a superconductor are electrically charged rather than neutral. This has a profound effect on the physics, which goes far beyond the physics of superconductivity. The fact that electronic charge is involved forces us to include a description of electromagnetic fields and electromagnetic particles known as photons.

Electromagnetism is known as a *gauge theory*. This is an expression that there are numerous mathematically equivalent ways to describe the electromagnetic field. This is exactly analogous to the existence of many languages. The underlying description of a tree is identical whether given in English, Japanese or Portuguese, but the sounds are quite different. When one breaks symmetry in a gauge theory the emergent particles are quite different from the previous examples. In fact the emergent particles disappear. They are eaten up by the particles of electromagnetism which themselves acquire mass (or grow fat). We are left with massive, emergent electromagnetic

particles in the system. This is the famous **Higgs mechanism**, which was first discovered in a superconductor [4].

- **Screening, asymptotic freedom and magnetic dirt in a metal**

We finish with an example of the use of the renormalization group. Remember that the renormalization group tells us how the "constants" of nature change as we change the length scale in which we're interested. Let's examine the charge of an electron. By carrying out the renormalization procedure we find that as we measure the charge at longer and longer distances from an electron the electric charge appears to get smaller. It tends to a limit as the length gets very long which is the small number that appears in all of the textbooks (which claim that the electron has a charge of $1.60217646 \times 10^{-19}$ C). How can this be? We were always taught that the electron has a fixed charge – not a number that changes depending on where you do an experiment! The answer is known as *screening*. What we think of as empty space is not at all empty; it is full of quantum mechanical particles buzzing in and out of existence. These particles may be thought of as owing their existence to Heisenberg's uncertainty relation $\Delta E \Delta t \approx \hbar$, which tells us that we may borrow enough energy ΔE to make a particle, as long as we only do so for a time Δt (and agree to return the energy afterwards!)

As a result an electron carries round with it a cloud of these short-lived "virtual" particles which themselves carry electric charge. The effect of these is to shield the apparent value of the electronic charge of the original electron. We conclude that the electron has an immense charge, far larger than the value quoted in the textbooks, but that the screening cloud carried around reduces it to the comparatively meagre value we measure in out experiments. However, if we can get close enough to the electron then we are able to penetrate the screening cloud and see a little more of the true charge of the electron. This explains physically the strange fact predicted by the renormalization group that the charge appears larger the closer we are to an electron.

Screening is nothing new. It has been known about in metals many years before this discovery. That is to say that an electron in a metal has its charge shielded from other electrons in the metal by a cloud of charges. However, this teaches us something profound: the emergent particle structure in solids is rather similar to that of the Universe. We think about the screened charge of an electron in a metal as somehow being artificial in that we may remove the electron from a solid and see its "true" properties. However, the properties of the electron in free space are also screened and it is impossible to see the "true" charge of the electron. In fact it is probably nonsensical to speak of these "true" values, which mathematically speaking, appear to be ill defined [5].

Finally, it's worth noting a counterexample of this screening effect, which represents one of the greatest triumphs of the renormalization group. There are some cases where, as we move further from a particle its charge appears to get larger! This *antiscreening* effect is known as **asymptotic freedom** [6]. It is seen in free space in the physics of quarks (the constituents of protons and neutrons). Quarks have a property known as "color charge", which is a little like electric charge and could be thought of as the reason that quarks in a nucleus are glued together. As we move away from a quark antiscreening causes the number of color charges build up, making the size of the colour charge seem larger and larger. Eventually this causes new particles to form, which are released in experiments as jets of matter, which may be measured in high-energy particle physics experiments.

Asymptotic freedom might seem like a very abstract idea whose application is only to the region deep inside atoms (and in exotic particle experiments taking place at places like CERN). However, it is also possible to find asymptotic freedom emerging in the condensed matter

physicist's laboratory. Like screening, the antiscreening associated with asymptotic freedom can take place in metals. Some metals are dirty: they contain atoms of another species inside them. If these atoms are magnetic they may attract an electron towards them with a force showing asymptotic freedom. As the electron moves further away, the force becomes larger and larger. The electron and the magnetic dirt are attracted to each other in the same way as quarks in a nucleus. This phenomenon, known as the **Kondo effect** (after Jun Kondo, who first explained its physics) [7], is almost exactly analogous to the asymptotic freedom of quarks and has been widely studied and discussed in condensed matter physics in the last fifty years.

Conclusions and outlook

This has been an attempt to introduce two fundamental concepts from condensed matter physics: broken symmetry and the renormalization group. These are unifying ideas that underpin condensed matter physics and show the emergent structure of the subject. Furthermore they also show how subjects from across physics may be unified into a picture of Nature involving emergent particles whose properties follow from the renormalization group. Perhaps the most attractive feature of this view of the world is that it may be tested, to high precision, in the laboratory both in exotic particle physics experiments but also, crucially, in condensed matter physics experiments.

In my talk I will review these ideas and show some results from a number of experiments, mostly taken from condensed matter physics, which are well described by this view of the world.

References

[1] P. W. Anderson, Basic Notions of Condensed Matter Physics (Westview Press, Colorado 1984).
[2] R. B. Laughlin, Nobel lecture, Rev. Mod. Phys. **71**, 863, (1999).
[3] S.J. Blundell, Magnetism in Condensed Matter (Oxford University Press, 2001).
[4] This description of the Higgs mechanism in terms of particles growing fat was invented by Sidney Coleman. See S. Coleman, Aspects of Symmetry (Cambridge University Press, 1988).
[5] The mathematical job of seeking meaningful values of these coupling constants such as charge and mass is known as *renormalization* and was a controversial subject for much of the 20th century. It is fair to say that the renormalization group has largely (albeit indirectly) solved this problem.
[6] For a gentle introduction to asymptotic freedom, see the Nobel lectures (by Gross, Politzer and Wilczek) for the 2004 Nobel Prize in Physics.
[7] P. Coleman, Introduction to Many Body Physics (Cambridge University Press, 2012).

Nov.27 (poster presentation)

P-01

'The Emergence of "Tipping Points" Games in Climate Change Debates: Scientific Versus Public Discourses'

Pojanath Bhatanacharoen
David Greatbatch
Timothy Clark
Durham Business School, Durham University

Summary

This paper compares and contrasts scientific versus public discourses in climate change discussions. It contributes to our understanding of communications strategies in public debates which impact on policy-shaping and, ultimately, the physical world. In particular, it explores the boundaries between scientific knowledge and rhetorics employed to motivate action. Using a discourse analysis approach, this paper presents empirical analysis on the nature of climate change discussions. In exploring the emergence and diffusion of the buzzword 'tipping point' in climate change debates, we firstly identify the thought leaders who wield influence in scientific and public communities. Secondly, we critically evaluate the discursive and textual variations between the scientific discourses (which try to explain the physical world) and the public discourses (which reflect the conceptual and political framing of climate change issues and thus shape our policy actions). The findings indicate that, while public discourses tend to draw upon scientific findings, the intellectual splits in academic communities render contentious the concept of 'climate change tipping points' as a term to describe the physical world.

Keywords: climate change, tipping points, diffusion, scientific discourse, public discourse

1. Introduction

The term 'tipping point' in its most basic meaning refers to a critical point when unprecedented changes occur rapidly with irreversible effect. It was first coined by the political scientist Morton Grodzins in 1957 in his studies on racial segregation to describe the critical threshold when the white population would leave an area where more and more black people were present. The 'tip point' is understood as a point where the increase in small and incremental changes have exceeded a certain threshold and 'tipped' the system out of balance leading to radical changes. The use of this term has spread across many disciplines ranging from sociology (Wolf, 1963; Stinchcombe et al, 1969; Pryor, 1971; Schwab and Marsh, 1980) to health sciences (Davis, 2000; Anderson, AF; Qingsi, Z; Hua, X, et al., 2003). Malcolm Gladwell, an American journalist, further popularized the term in his book 'The Tipping Point: How little things can make a big difference' (2000).

Most notably, 'tipping point' is often used in climate change contexts by climatologists, environmental activists and politicians (Russill 2008). A LexisNexis database search revealed that approximately 8.7% of all 'tipping point' articles in English language newspapers are related to climate change, indicating a high frequency of media attention in this area. As Walker (2006; 802) commented: 'In 2004, 45 newspaper articles mentioned a "tipping point" in the connection with climate change; in the first five months of this year [2006], 234 such articles were published. "Warning hits tipping point", one UK newspaper recently warned on its front page; "Climate nears point of no return," asserted another. The idea is spreading like a contagion.'

Yet, as a scientific concept to describe a threshold that depended on anthropogenic factors, it is not innovative, and there are alternative terms that can be used. Against this background, this paper examines the usages of the 'tipping point' concept in the climate change contexts. It considers: how concepts such as 'tipping point' spread in academic and popular domains; whether shifts in scientific

discourse lead to a shift in public discourse and vice versa; and the consequences of using a term such as 'tipping point' on policy debates..

2. Methodology

We used a combination of citation, content and discourse analysis to trace the usage and interpretations of the tipping point concept in academic journals and newspaper articles. We searched the ISI Web of Knowledge database of major academic journals using the search term 'tipping point' and then categorized the data into different disciplines. We also searched the LexisNexis database which covers the major newspapers for English language news articles using 'climate change' and 'tipping point' combined as a search term.

3. The emergence of the tipping point metaphor

According to our ISI Web of Knowledge search, Lindsay and Zhang (2005) is the first academic article on climate change which uses the term tipping point in its title and abstract. However, Lindsay indicated in an interview that he could not recall where he came across the term and that it was not central to the paper.

Our LexisNexis search indicated that the first time the term 'tipping point' was used in the climate change context in non-academic media was as early as in 1998 (10 December 1998, The Atlanta Journal and Constitution).

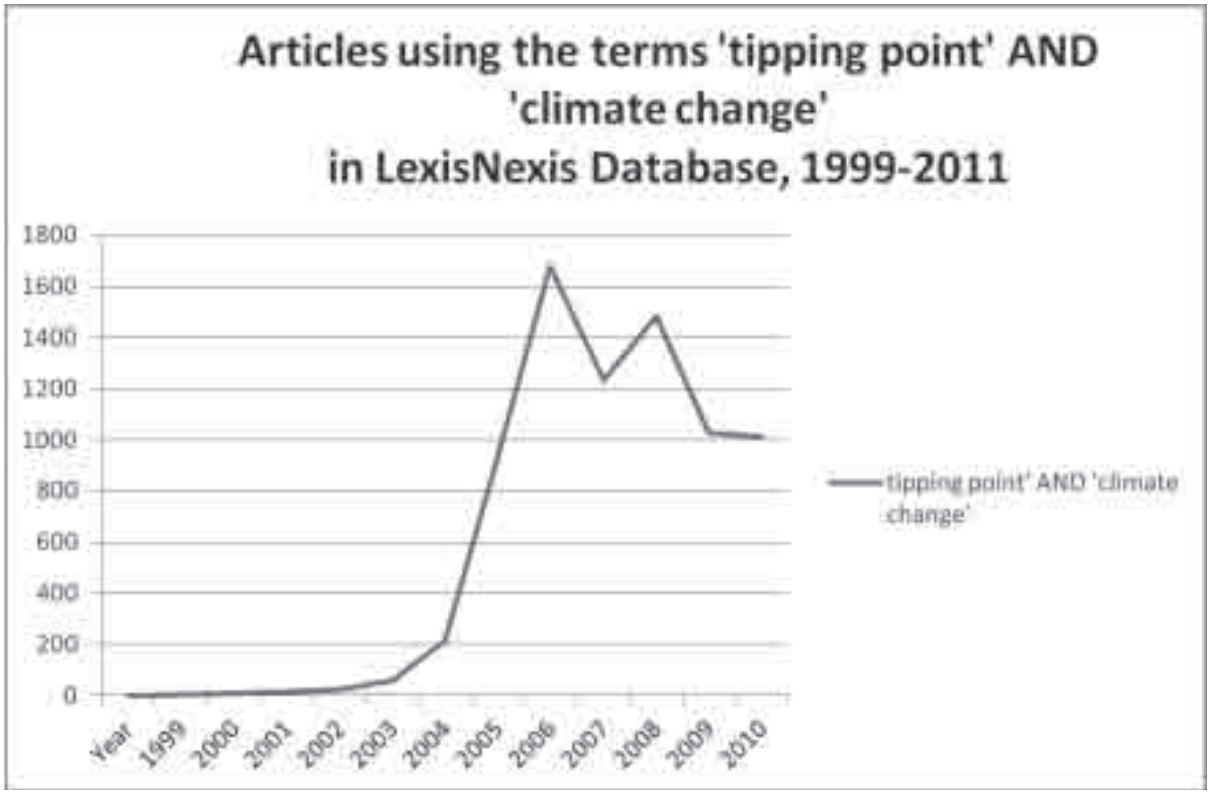
> Even confirmation that the Earth's warming this century is unprecedented for at least the previous 11 centuries does not prove that global warming has become a reality. But people such as Jonathan Overpeck, head of the paleoclimatology program for the National Oceanic and Atmospheric Administration, are more than a little worried.
>
> "There is no period that we can recognize in the past 1,200 years that was as warm on a global basis," Overpeck told the annual meeting of the American Geophysical Union this week. "That makes what we're now seeing more unusual and more difficult to explain without turning to a 'greenhouse gas' mechanism."
>
> Other scientists warn that if global warming is truly occurring, the somewhat subtle changes we are seeing today aren't our real worry. They point out that complex systems such as climate tend to stay relatively stable for a long time, absorbing pressure for change until it reaches a tipping point. Once that point is reached, major change occurs quickly, in a relative snap of the fingers. Such rapid and abrupt climate change would make it difficult for plant and animal species to adapt, particularly when many of those species are already under stress, thanks to other man-made pressures.

One individual was mentioned in this article, the climatologist Jonathan Overpeck. A search of his work reveals that he does not use the 'tipping point' but adopts the concept of 'abrupt climate change' in his research. Also, the article indicated that '[o]ther scientists' were using the 'tipping point' concept but their names were unspecified.

What does it mean in terms of our understanding of how ideas diffuse? First, once an idea is in the public domain, it becomes difficult to pinpoint exactly how it cascaded from one individual or context to another. Second, the original thought leaders are not always the ones responsible for the popularization of an idea and may not be readily identifiable. The usage of the tipping point concept in the climate change context seems to have originated from a scientific community, first appearing in the media in 1998 before Gladwell's book popularized the term in 2000. It is possible that the unnamed scientists picked up the term from Gladwell's 1996 New Yorker article but we do not know for certain. Nevertheless, as figure 1 shows, the usage of the term tipping point in the climate change context did not rocket until 2005 which indicates that neither the scientists cited in the 1998 newspaper article nor Gladwell may have been responsible for the exponential growth of the tipping point usage in the climate change context. On the other hand, Russill (2008) indicates that James

Hansen's (a NASA scientist and environmental activist) "tipping point" forewarnings received prominent media coverage in early 2006 (134) (1752-1753).

Figure 1 Articles using the term 'tipping point' in the LexisNexis database



4. **The use of the tipping point metaphor**

In academic articles the 'tipping point' metaphor is used to describe physical processes in a relatively neutral way. Thus, for example, Lindsay and Zhang (2005; 4879) indirectly defined the tipping point as a point where the accumulation of small changes leads to a big topical change in a system: 'The late 1980s and early 1990s could be considered a tipping point during which the ice-ocean system began to enter a new era of thinning ice and increasing summer open water because of positive feedbacks. It remains to be seen if this era will persist or if a sustained cooling period can reverse the process.' Lindsay and Zhang use of the tipping point concept is not filled with epidemiological imaginary. Moreover, they carefully pointed out that it remained to be seen whether the process can be reversed, and did not paint an apocalyptic picture. Winton (2006) too did not use epidemiological imaginary: 'If a glass is slowly tipped with a finger, it eventually reaches a point where its upright equilibrium becomes unstable and it proceeds rapidly to a new stable equilibrium on its side (2006; 1).'

Nevertheless, 'tipping point' is a contested term and there is a split in the scientific community as regards the appropriateness of using it in climate change discussions. Kenneth Caldeira, an earth scientist, suggested in a New York Times article by Revkin (2009) that 'I think a lot of this threshold and tipping point talk is dangerous... If we say we passed thresholds and tipping points today, this will be an excuse for inaction tomorrow...' There is a concern for potential political inertia. Gardiner (2009) observes that there are recent shifts in climate change debates from the 'gradualist approach' in which climate change occurs gradually and in a linear fashion to the 'abrupt paradigm' which operates on the premise that there is an anthropogenic climate threshold which once crossed will lead to

catastrophic consequences. While Gardiner suggests that the alarmist approach reflects the activists' efforts in stirring climate change actions, he argues that the 'abrupt change paradigm' does not lead to significant policy changes. Notably, he observes that, despite the threat posed by 'abrupt climate change', the progress in reaching political agreements for issues such as CO2 admission has been minimal and global agreements have been weak and violated. Moreover, Gardiner argues that the 'abrupt climate paradigm' worsens the political inertia as the impending temporal proximity of apocalyptic scenarios leads to fatalism in which people resign to the catastrophic fate rather than resist it. This raises the questions of why 'alarmist' concepts such as 'tipping point' are still being used and what are the links between the scientific and public 'tipping point' discourses and policy formation.

Russill suggested that 'tipping points' were imbued with epidemiological references which evoke a sense of viral contagion with techniques and styles of reasoning which capture wider public imagination. He concluded that 'Climate systems might be described in a variety of manners, either using tipping points or other concepts. Its importance derives not from better describing climate systems but in making available an image of crisis useful for registering public concern and opening avenues for response (146-147).'

Hansen first applied the tipping point metaphor at an address to the American Geophysical Union. He said: 'I present multiple lines of evidence indicating that the Earth's climate is nearing, but not passed, a tipping point, beyond which it will be impossible to avoid climate change with far ranging undesirable consequences.' Although Hansen did not use epidemiological imaginary, he did indeed use the tipping point metaphor not merely to describe the climate problem to a public audience, but to raise public and grassroots awareness and action. This is further demonstrated in his following statement:

> There is little merit in casting blame of inaction, unless it helps point toward a solution. It seems to me that special interests have been a roadblock wielding undue influence over policy-makers. The special interests seek to maintain short-term profits with little regard to either the long-term impact on the planet that will be inherited by our children and grandchildren or the long-term economic well-being of our country. The public, if well-informed, has the ability to override the influence of special interests, and the public has shown that they feel a stewardship toward the earth and all of its inhabitants. Scientists can play a useful role if they help communicate the climate change story to the public in a credible understandable fashion.

This is a rhetorical and direct plea to the public audience to act against special interests for the greater good. It is noteworthy that although Hansen is a scientist, his usage of the tipping point term did not first appear in his academic writing, but in the wider public sphere. This is consistent with Russill's (2008) suggestions that scientists and politicians have used the tipping point metaphor on its more apocalyptic end of its meaning spectrum to attract public attention and motivate public action.

Climatologists Lenton et al (2008; 1786) indicated that 'Many of the systems we consider do not yet have convincingly established tipping points. Nevertheless, increasing political demand to define and justify binding temperature targets, as well as wider societal interest in non-linear climate changes, makes it timely to review potential tipping elements in the climate system under anthropogenic forcing.' In other words, scientists have not been able to conclusively demonstrate whether many aspects of climate change are in fact reaching a tipping point. Yet, there is an urge to influence policy-making, and, therefore, scientists should define parameters under which climate change problems can be framed through a policy lens.

Bibliography

Anderson AF, Qingsi Z, Hua X, Jianfeng B. (2003). 'China's floating population and the potential for HIV transmission: a social-behavioural perspective' *AIDS Care-Psychological and Socio-Medical Aspects of AIDS/HIV,* 15/2: 177-185.

Davis, RM. (2000) ‘Moving tobacco control beyond "the tipping point" - Ample funding, strong policies, and "unsticky" cigarettes are key’ *British Medical Journal,* 321/7257: 309-310.

Gardiner, Stephen, “Saved by Disaster? Abrupt Climate Change, Political Inertia, and the Possibility of an Intergenerational Arms Race,” *Journal of Social Philosophy* 40(2) 2009, 140-162.

Gladwell, M. (2000) *The tipping point : how little things can make a big difference***.** London : Abacus.

Grodzins, M. (1957) ‘Metropolitan Segregation’ *Scientific American*, 197/4: 33-41.

Hansen, J. (2005) ‘Is there still time to avoid “Dangerous Anthropogenic Interference” with Global Climate’ A tribute to Charles David Keeling. Presentation given December 6, 2005 at the American Geophysical Union (http://www.columbia.edu/~jeh1/2005/Keeling_20051206.pdf)

Lenton, TM; Held, H; Kriegler, E, et al. (2008) ‘Tipping elements in the Earth's climate system’ *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 105/6**:** 1786-1793**.**

Lindsay, RW; Zhang, J. (2005) ‘The thinning of Arctic sea ice, 1988-2003: Have we passed a tipping point?’ *Journal of Climate*, 18/22: 4879-4894.

Pryor FL. (1971) ‘Empirical Note on Tipping Point’ *Land Economics*, 47/4: 413-417.

Revkin, Andrew, March 28, 2009. Among Climate Scientists, a Dispute Over ‘Tipping Points’ http://www.nytimes.com/2009/03/29/weekinreview/29revkin.html?_r=1#articleBodyLink

Russill, C (**2008)** ‘Tipping Point Forewarnings in Climate Change Communication: Some Implications of an Emerging Trend’ *ENVIRONMENTAL COMMUNICATION-A JOURNAL OF NATURE AND CULTURE* 2/2**:** 133-153**.**

Schwab, WA; Marsh, E. (1980) ‘The Tipping-Point Model – Prediction of change in the Racial Composition of Cleveland, Ohio, Neighborhoods, 1940-1970’. *Environment and Planning A*, 12/4: 385-398.

Stinchcombe, AL; AcDill, M; Walker D. (1969) ‘Is there a Racial Tipping Point in Changing Schools’. *Journal of Social Issues*, 25/1: 127-136.

Walker, G (2006) ‘The tipping point of the iceberg’ *NATURE* 441/7095**:** 802-805**.**

Winton, M (2006) **‘**Does the Arctic sea ice have a tipping point?’ *GEOPHYSICAL RESEARCH LETTERS* 33/23**.**

Wolf, EP. (1963) ‘The Tipping point in Racially Changing Neighborhoods’ *Journal of the American Institute of Planners*, 29/3: 217-222.

P-02

Beauty as protection against environmental instability

James Woodward
Durham University

Summary

In my presentation I will discuss the importance of beauty with regard to the way in which we interact with our environment as everyday citizens but also as scientists and academics. I will argue that it is vital that beauty is seen in the correct light, and that it is a matter of metaphysics and not a simple matter of taste. I will further suggest that we should consider that sense and reason are fundamentally interrelated and one cannot be given more significance than the other if we are to be truly human, and to responsibly consider our place on Earth. To facilitate this I will examine the view that in the post-Enlightenment age we have adopted a way of seeing the world which places reason as both superior to and abstracted from sense. Furthermore, I argue that this bifurcation has dangerous implications for our environment.

As a way of moving forward, I commend an aesthetic education, as conceived of by Friedrich Schiller, to be an excellent way of improving ourselves morally. I will briefly discuss Schiller's analysis of the human condition demonstrating that by allowing both the rational and sensual aspects of ourselves to flourish, we can function on a more profound level. This mode of life, which places the beautiful in a prominent position gives us the chance to be more skilled morally. With the scientific and the poetic working together, we are appropriately equipped to make the best decisions for our environment so that both we and future generations may benefit.

Introduction:

At first glance, the title appears to refer to the construction of aesthetically appealing objects in our world. It is perhaps not immediately obvious how something such as beauty could have any affect on the environment in a pragmatic sense. I will argue, however, that beauty is fundamental to our moral being and that a deep understanding of this is essential to safeguard our environment.

In order to do this I will consider three things:
i) Beauty as the foundation for our moral being
ii) The degeneration of beauty and the triumph of reason
iii) A need to rehabilitate beauty by way of an aesthetic education

i)
When understanding beauty a vital distinction should be made: that of beauty *qua* taste and beauty *qua* metaphysics. It is the latter that I believe to be important and I shall illustrate this with some examples of what it is not: [a] that meal was beautiful; [b] that woman is beautiful; [c] that mountain is beautiful, and [d] that piece of music is beautiful. What [a] claims cannot be so, the meal may have been delicious but this is a matter of taste; in [b] this may well be so but it is likely that the reference is made as a matter of desire; with [c] it is slightly trickier but the 18th and 19th century use of 'sublime' is probably more accurate here; and concerning [d] this also may be the case but it is likely that this is in fact a judgment of taste.

We should consider this distinction carefully for otherwise we would subsume things that are appealing to our senses, things that we are in awe of, or things we desire under the same term. When we consider that the beautiful and the good may be seen as intrinsically linked, then we surely would not want to suggest that what we deem to be morally good should fall under the same evaluative process as what songs make us want to dance, what we like to eat or who we would like to take to supper.

The importance of the distinction between beauty as a matter of taste or of metaphysics can be traced to Plato. The difficulty with sense based judgements, such as taste, is that our senses can be deceived. We only have to consider eye witness testimonies to discover how different persons can record the details of an event in contradictory ways. Plato argued that there are two different realms, the visible and the intelligible and it is to the first realm that matters of taste are affiliated. The visible realm is where matters of opinion find themselves but are essentially, in this view, shadows and images. The intelligible realm is where matters of truth can be located, and these are found through mathematical reasoning and intelligence. It is here also where Plato suggested that the Forms are, the archetypal structures behind our world, including the Form of the Good. This is of importance to us because the beautiful and the good are seen to be intimately related - in *The Symposium*, Plato suggests that the love of the beautiful would lead us to the Good (love requires beautiful things, good things are beautiful, therefore love requires good things).

Consider Plato's warning that art and beauty should be kept firmly apart in light of this judgment. Art is considered to be sense based representation (in fact the representation of a representation), and so is illusory and likely to lead us away from the truth. Keeping art and beauty apart ensures that the Good, the highest Form and value, is not tainted by falsehood. Following this then, it would make no sense whatsoever to say that a meal etc. was beautiful.

ii)

It is often easier to say what something is not than what it is and beauty is no exception. In this section, I will consider that beauty and nobility are very similar (with the understanding that the beautiful and the good may be linked). If we see someone acting nobly, the athletes at the Olympics showing composure under pressure or even assisting their fellow athletes as examples, we are inspired by their nobility, by their beautiful behaviour. When we consider the environment, it seems that most people are inspired by a sunrise or a sunset (a visit to an art gallery will see the amount of photographic images and paintings dedicated to such matters). I am sure there are many reasons for this but am convinced that this is due to the nobility of the sun and the stillness and grace with which the sun rises and brings new possibility to the day; juxtaposed with the quiet tranquility as the sun slowly sets and brings our day to an end. Simply put, we revere the noble and are protective towards the beautiful, and it is a inherent aspect of our humanity, or at least it should be.

Part of the reason we pollute our environment may well be an aspect of how we externalise feelings about our lives, but this is beyond the scope of this discussion. What I argue, however, is that there is a metaphysical assumption and precedent for how we view and interact with our environment. If we consider the recent events surrounding Fukushima, we can see that there are two conflicting views as to how science and technology relate to our world; one being that science and technology (in this case

nuclear power) are bad and irresponsible, the other that this technology is largely safe, good and important for us to sustain ourselves.

While we could easily spend the whole of this presentation and more considering what metaphysics entails, the definition I offer is this: the paradigmatic explanation of what we perceive [consider] to be reality. A difficulty well considered in hermeneutics is that we can never truly be objective about our world. As Camus remarked:

> 'Understanding the world for a man is reducing it to the human, stamping it with his seal. The cat's universe is not the universe of the ant-hill. The truism "All thought is anthropomorphic" has no other meaning.' [Myth of Sisyphus]

This is not to say that there is only humanity; we should be mindful that to be anthropomorphic does not necessarily imply being anthropocentric. This leads us to consider how, paradigmatically, we place ourselves in the world and how we interact with it. Heidegger offers an excellent analysis as to how this view has changed over time through use of the terms *mythos* and *logos*, the former relating to myth and poetry, the latter to naming and reason. Heidegger noted that the Greeks viewed the world by way of *mythos*, and this view made *techne* (art/skill/craft) connected with *physis* (nature), so man would interact with nature harmoniously. The example given is bread making, i.e. to produce what nature does not originally offer. In modernity, we view the world by way of *logos,* and according to this view *techne* and *physis* are taken apart. Man does not seek to work in harmony with nature, rather he seeks to control it.

The change from *mythos* to *logos* occurs as the development of modernity took shape. When Nietzsche warned of this at the end of the 19th century he was ridiculed but his warning is still worth our attention:

> 'Are we not straying as through an infinite nothing? Do we not feel the breath of empty space? Has it not become colder? Is not night continually closing in on us? Do we not need to light lanterns in the morning? Do we hear nothing as yet of the noise of the gravediggers who are burying God? Do we smell nothing as yet of the divine decomposition? Gods, too, decompose. God is dead. God remains dead. And we have killed him.' [GS §125]

This passage, as with many of Nietzsche's, has been often read out of context. The argument here is that God is actually dead, rather that the paradigmatic way of viewing the world through a God-shaped lens has decayed. His warning is to all of those who rejoiced at the birth of modern science and the apparent rational way in which the world could now be understood, without recourse to the superstition and primitive understanding that was inherent in former times. The warning is that in destroying this way of seeing the world, we create a [metaphysical] vacuum that needs to be filled. The world as it is (as Kant so wisely suggested) is beyond our understanding and remains so regardless of our fanciful explanatory systems.

The problem is that while humanity was using its skills to work in harmony with nature, to see an interrelation with the world was assumed. Mythological and religious tales secured this understanding,

that acting well was important and if this duty was neglected then we would be punished by the gods. In that situation, human beings were a constituent part of the universe and certainly not at the centre. With the hubris of modernity, the view becomes staunchly anthropocentric and human ends are seen to be the ends of the world, that our development and advancement must be inherently good. From this we equate technological advancement with what is good.

My assessment, however, is that we cannot see technological advancement as good without qualification. As human beings we do not possess the ability to understand the impact of our involvement with the environment and we could be bequeathing a mess to future generations. I am certainly not arguing for regression, I am in awe of many of the things that science can do for us. The plane that flew me here is a stunning example of human discovery, the engineering behind many of the buildings in Tokyo enabling them to bend and survive earthquakes is quite tremendous. Technology should not be seen as a *de facto* good and limits [ethics] should be firmly placed upon its advance. While I accept the arguments that the situation in Fukushima could have been a good deal worse, that it was already an old power station, and technological advances have made nuclear power safer than ever, this does not mean that we should accept it without question. We have seen the effect of technology utilised for war on an industrial scale in 1914-1918; the initial research into nuclear physics led to the creation of atom bombs.

Aristotle said that we cannot judge the moral value of a man until after his death (for many reasons including the results of actions coming to light), and none of us present here perhaps are able judge the value of nuclear power. The paradigmatic aspect of our epoch remains that which we were bequeathed by the Enlightenment, i.e. that reason is the core aspect of humanity and to its dictates we should ascribe. What I argue in the next section is that reason is indeed important but should not be placed above the sensual in understanding both human beings and how we are to be in the world.

iii)

The idea of beauty being of significance when referring to the environment seems, in philosophical terms, a thin rather than a thick suggestion. The idea that pragmatism is to be preferred over romanticism, that we should act according to a firm notion of reality rather that a romantic ideal of how the world could be, may be what one considers when first reading this. This facile reading of beauty, however, relates to the sense of beauty *qua* taste (thin) rather than the beauty *qua* metaphysics (thick). When we consider beauty *qua* metaphysics, it makes sense to consider that working to develop our understanding and appreciation of beauty may have a moral impact upon our lives.

This assertion has firm foundation in the thought of Friedrich Schiller, who argued that we should seek progress by way of an aesthetic education. This is not to be understood by as a History of Art course or a way of learning to appreciate fine art, as this would conflate thin and thick readings of beauty in favour of the former. In developing our aesthetic sensibilities, our moral sensibilities may benefit. For example by learning a musical instrument we may learn to deal with frustration and this can teach us the importance of patience when dealing with irascible people. If we learn to be more kind and considerate to the shortcomings of others, we may be able to evaluate things less judgmentally.

The greatest strength of Schiller's analysis, however, is that he takes into account both the rational and sensual aspects of mankind and suggests that we need to satisfy both parts. The form and the sense drives, in Schiller's terms, are held to be our key motivational aspects. The problem faced by philosophers preceding Schiller was to see one or the other of these being primary [or in the ascendancy], for example by way of rationalism vs romanticism.

Schiller suggests, however, that there is a third drive. The play drive is found when the two other drives harmonise, and this creates a deeper level of involvement. If we consider fire and water, it does not make sense to say that one can exist without the other, nor that one is of more value than the other. If they are combined in the right way, they create steam and this is of a higher order. Or in a slightly more poetic formulation, which expresses the need for this higher order, Nietzsche indicates how easily someone could engage with a complex world after receiving the appropriate training:

Glattes Eis	Smooth ice
Ein Paradeis	is paradise
Für den, der gut zu tanzen weiss	for those who dance with grace

Nietzsche [GS prelude §13. For Dancers]

In my discussion, I have argued that we need both reason and sense, both science and poetry, when attempting to understand our world and our place in it. I have also asserted that beauty is a matter for metaphysics and as such is interwoven with morality. A greater sensitivity to the beautiful can lead us to a more harmonious and ethical approach to our environment. It is in beauty that we find expression of our nature as human beings. As noble humans who act according to the principles of beauty, we can improve our lives and the lives of others, and that includes the future generations who will inherit what we leave to them today.

P-03

Emergence of Conceptual Foundation Blocks of Chemistry from Fundamental Quantum Mechanics

P. R. Tulip
Department of Physics
University of Durham
UK

Summary: Chemistry provides a conceptually powerful and intuitive description of matter comprised of 'building blocks' such as atoms, orbitals and spins. However, this picture relies upon being able to unambiguously partition and define the properties of a system into discrete 'chunks'. Such a partitioning is problematic in condensed phase systems (i.e. solids or molecules), and the connection to the underlying many-body quantum mechanics that governs the behavior of such systems is opaque. In this presentation, I will discuss the reasons for this, and discuss how modern treatments of the many-body quantum mechanical problem such as density functional theory (DFT) lead to emergent properties that may be interpreted within the framework of the traditional chemistry perspective.

1. **Introduction**: The "molecular hypothesis" has provided an intellectually satisfying and coherent model with which to interpret, understand, and to predict the chemical properties of systems. At its heart, this model is based upon the idea that one can unambiguously ascribe particular properties to molecules and fragments of molecules such as functional groups. One also invokes concepts such as electronic orbitals to understand the chemical behaviour observed, where an orbital is a single-particle wavefunction that describes the behaviour of an electron in a molecule or atom. Single-particle objects such as this are conceptually 'easy' to envisage; they align well with our classical intuition and prejudices that we can distinguish and label distinct particles. However, if one wishes to understand the properties of a many-particle quantum mechanical system, the appropriate (non-relativistic) starting-point is the Schroedinger equation:

$$H\Psi = E\Psi$$

where H is the Hamiltonian governing the dynamics of the system.

The many-body wavefunction Ψ is normalized to one (i.e. its modulus squared integrates to one), and thus can be interpreted as a probability amplitude for finding N particles at locations $\{\mathbf{r}\}$. This is a mathematically complicated object that cannot be decomposed into single-particle contributions. Thus, at a fundamental quantum mechanical perspective, it becomes impossible to talk about single-particle properties; instead, we must consider the many-body wavefunction, and discuss properties of the system as a whole.

Unfortunately, the many-electron Schroedinger equation is intractable to analytic solution, i.e. it is not possible to derive a closed-form solution using pen and paper.

Instead, one must endeavour to solve this equation numerically, that is, to use modern computers to grind through the myriad calculations required to determine a solution. This is a formidably difficult task!

The requirement that we describe the electrons quantum mechanically has other implications, too: as we are now working with a wavefunction describing a probability distribution, classical concepts such as the electron possessing a well-defined location or trajectory must be abandoned. We must also accept that these are indistinguishable particles. This has profound implications: we can no longer speak of certain electrons 'belonging' to given atoms, and consequently, the very concept of 'what is an atom in a molecule or solid?' becomes ill-defined; a perhaps surprising conclusion. We appear therefore, to have two distinct pictures of reality: a mathematical formulation of many-body quantum mechanics, mathematically and physically rigorous, yet formidably difficult to solve, and non-intuitive to interpret, and a phenomenological 'chemical' picture, in which we view the world as being comprised of 'intuitive' building blocks such as atoms, molecular fragments and molecules, each with their own distinct character, and where the properties of a whole may be understood in terms of these blocks and the 'bonds', comprising electronic orbitals, that form between them. The connection between these two pictures of reality seems, at first sight to be opaque; indeed, it would seem, from the description given above, that the two approaches fundamentally disagree.

This presentation will discuss how we can resolve this discrepancy using modern treatments of the many-body quantum mechanical problem.

P-04

Direct Numerical Simulation of Electrophoresis of Charge Colloids in AC field

ChunYu Shih

Kyoto University

Many researchers investigate electroacoustics phenomena using both theoretical approaches and experiment techniques to calculate the mobility of high concentration suspension. It reviews that electroacoustics is a powerful tool to quantifying colloidal systems at steady state. The transient behavior of these phenomena is also important as it provide insight of relaxation time and polarization effect. In this work, we carried out direct numerical simulation (DNS) of dynamic electrophoretic mobility of charged colloidal dispersions in an oscillatory electric field using smooth profile (SP) method. In this method, we fully take into account electro-hydrodynamic couplings by solving Newton, advection-diffusion, and Navier-Stokes equations. This enabled us to compute the time evolutions of colloidal particles, ions and host fluids simultaneously. The electrophoretic mobility of dilute dispersion can be explained well by approximation theories at low frequency, but deviates in the case of dense dispersions. We also pay close attention to the electric double layer deformation. The results show that the deformation of electric double layers is the main factor of polarization effect.

1. Introduction

Electroacoustic is a phenomenon coupling between ultrasound and electric field, which can be divided into two basic types: (1) when acoustic pressure propagates through a colloidal suspension or emulsion, it creates a macroscopic potential difference so-called the colloid vibration potential (CVP); and (2) the obverse effect when an alternating electric field propagates through a colloidal suspension or emulsion and this generates an acoustic field so-called the electrokinetic sonic amplitude (ESA). In both case, both theoretical theory and experimental techniques make electroacoustics device a powerful analytic tool in quantifying a suspension system. Because mobility of high concentration suspension is difficult to measure, it is the advantage of electroacoustics which collects the acoustic field when an electric field propagates through a colloidal suspension or emulsion instead of optical measurements. In this way, properties of concentrated simple such as PH, electrolyte concentration will not change without dilute, so we can get a more realistic response in high concentration suspension.

Here we introduced a method called the smoothed profile (SP) method, which can compute the time evolutions of colloidal particles, ions, and host fluids simultaneously by solving Newton, advection-diffusion, and Navier-Stokes equations so that the electrohydrodynamic couplings can be fully taken into account. The comparisons with approximation theories show quantitative agreements for dilute dispersions without any empirical parameters; however, our simulation predicts notable deviations in the case of dense dispersions.

2. Discussion

In order to overcome the problems arising at the solid-fluid interface, a smoothed profile was introduced to the interface. In this method, the surface of the colloid is treated not as a sharp interface having no thickness, but rather, an interface is introduced having a width comparable to the grid spacing. The colloid density profile is defined so as to change smoothly within this finite thickness interface.

In order to determine regions in which a colloidal particle exists, the density field is introduced:

$$\phi(r,t) = \sum_{i=1}^{N_p} \phi_i(r,t)$$

Here, $\phi_i \in [0, 1]$ is the density profile of the *i* th particle, which is unity in the particle domain $|\boldsymbol{r}| < a - \zeta/2$, zero in the fluid domain $|\boldsymbol{r}| > a - \zeta/2$, and has a continuous diffuse interface within the thin interface domain $a - \zeta/2 < |\boldsymbol{r}| < a + \zeta/2$ whose thickness is ξ. We define the special distribution of the surface charge $eq(\boldsymbol{r}) = Ze\,|\nabla\phi|\,/\,4\pi a^2$ using ϕ; then the local density of the total charge is represented smoothly everywhere in the system by $\rho_e(\boldsymbol{r}) \equiv \Sigma\alpha\, Z\alpha\, e\, C\alpha + eq$. The complete dynamics of the system is obtained by solving the following time evolution equations.

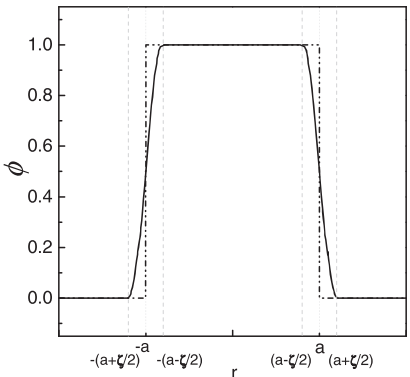


FIG. 1. An example of the smoothed profile method.

We have performed simulations for electrophoresis of a single particle and compared them with the O'Brine-White theory. A constant uniform DC electric field E = 0.1, which corresponds to 2.85 × 10 3 V/cm, was applied. A relationship between the dimensionless mobility $E_m \equiv 3e\eta V\,/2\epsilon k_B TE$ and dimensionless zeta potential with $\kappa a = 1.0$ is show in Fig 2(a). It is clearly demonstrated that our method reproduced the O`Brien-White theory almost perfectly including the non linear regime $y \geq 2$ within only a few percent error. Here the dimensionless zeta potential $y \equiv e\zeta/k_B T$ is introduced. Then we simulate a single particle in AC field. In Fig 2(b), the dimensionless mobility is plotted as a function of dimensionless frequency. We found that the magnitude of mobility has a local maximum, it is because the polarization. For the case when a static electric field is applied, double layer polarization will induce an internal electric field, which is in the inverse direction as that of the averaged electric field, and has the effect of retarding the movement of a particle (Fig 2(c)).

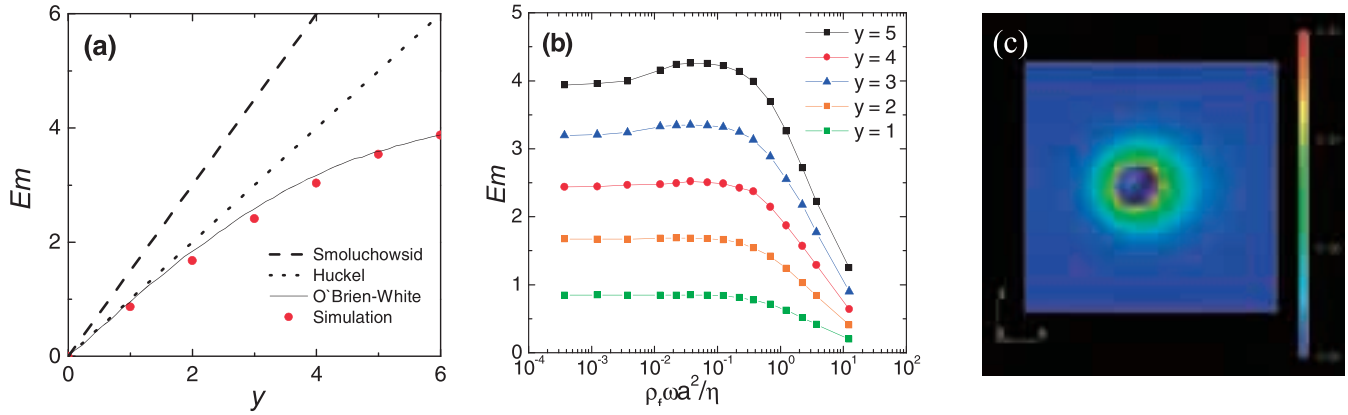
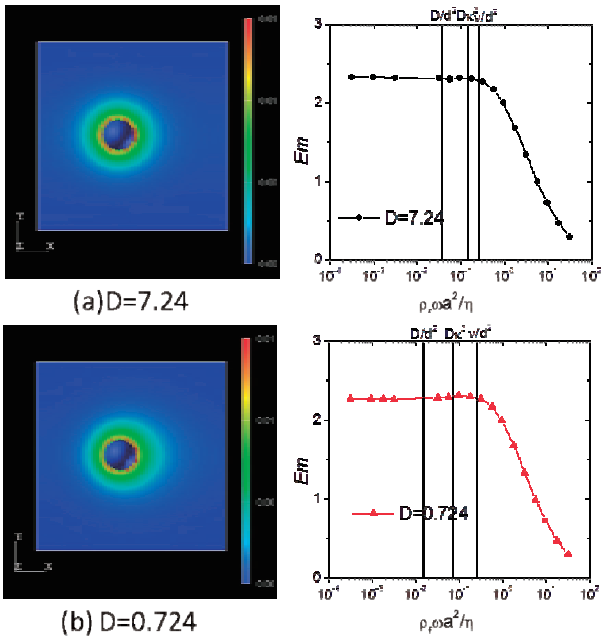


FIG 2.(a) Relationship between dimensionless mobility and dimensionless zeta potential.
(b) Relationship between dimensionless mobility and dimensionless frequency.
(c) Snapshot of colloid particle and double layer.

The effect of diffusion coefficient

In this case we change the diffusion coefficient D from 7.24 to 0.00724. Because of relatively slow diffusion of ions, the polarization effects become apparent. The distribution of charge density due to counter ions and co-ions are shown in Fig 3. It is quite evident that the electric double layer is deformed considerably in the small diffusion coefficient Fig 3(c), while it is almost isotropic in high diffusion coefficient 3(a).



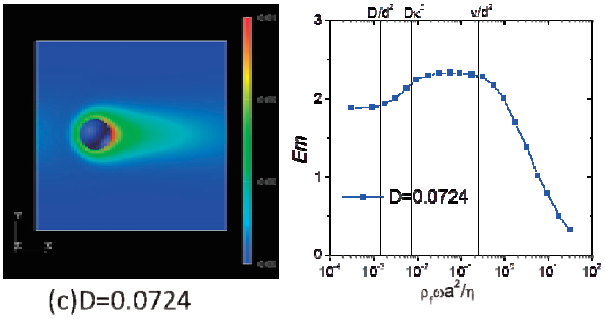


FIG 3. Relationship between dimensionless mobility and dimensionless frequency in AC field. The snapshot is the same electric field in DC.

The effect of volume fraction φ

Our simulation method is easily applicable to dense dispersions consisting of many particles. We thus examined the effect of the particle concentration on the electrophoretic mobility. Figure 4 shows typical snapshots of the systems with (a) fcc, (b) bcc, and (c) random configurations

Figure 4(d) shows the variation of the magnitude of the dimensionless mobility as a function of dimensionless frequency at various volume fraction of particle φ. As can be seen from Figure 4(d), if φ is low, the magnitude of the mobility decrease with the increase in H. This is due to the fact that the interaction between neighboring particles has the effect of retarding their movement.

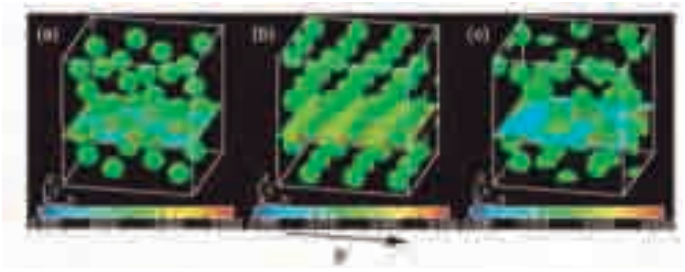
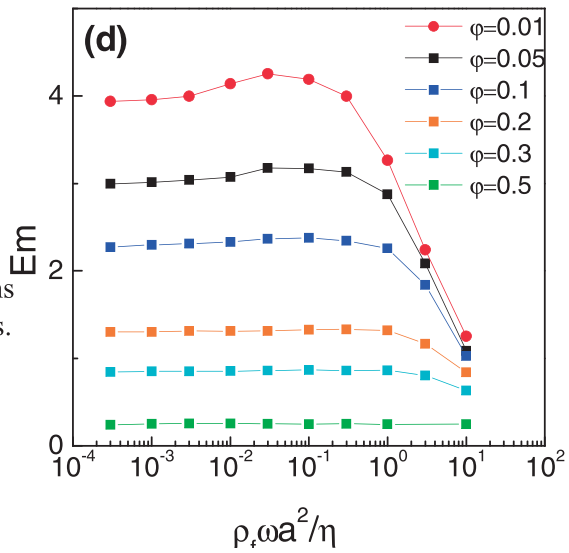


FIG 4. Snapshots of the electrophoresis of dense dispersions with (a) fcc, (b) bcc, and (c) random particle configurations. (d) is the mobility in different volume fraction φ.



P-05

Hydrodynamic Interactions in a Compressible Fluid

Rei Tatsumi
Department of Chemical Engineering, Kyoto University

Summary: We present a direct numerical simulation study for the dynamics of dispersed particles in a compressible solvent fluid. The hydrodynamic interactions play an important role in the dynamics of particle dispersions and give many attractive properties. The hydrodynamic interactions are the momentum transmission in a fluid by two forms as viscous diffusion and sound propagation. However, in many dispersion systems, the sound propagation is quite faster than the viscous diffusion, and the influence of sound mode is usually ignored by the assumption of an incompressible fluid. In this study, we investigated the role of the sound propagation in the hydrodynamic interactions.

1. Introduction: Particle dispersions have various unique properties, and widely used in our daily commodities such as paints, cosmetics, foods, and so on. Therefore, an understanding of these properties is important in many fields of science and engineering. These properties originate from the dynamics of particles, which are extremely complicated because of the hydrodynamic interactions among particles mediated by the motion of the surrounding fluid. When one particle in the fluid moves, a part of momentum of this particle is transferred through the fluid as a flow and affects on the motions of particles in the fluid. Such momentum exchanges are called hydrodynamic interactions.

The hydrodynamic interactions are transmitted in two ways: via viscous momentum diffusion and via sound propagation. Each of them has different time scale of propagation. The time scale of viscous diffusion over the particle size is $\tau_v = a^2/v$, and that of sound propagation is $\tau_c = a/c$, where a is the particle radius, v is the kinematic viscosity, and c is the speed of sound in the fluid. Therefore, the compressible factor is defined as the ratio of the two time scales as $\varepsilon = \tau_c/\tau_v = v/ac$. For many dispersion systems, this factor is quite small because of the high sound speed. For example, in the case of a dispersed particle of radius a = 100 nm in water, the compressibility factor is evaluated from $v = 1.0\times10^{-6}$ m^2/s and $c = 1.5\times10^{3}$ m/s to be $\varepsilon = 6.7\times10^{-3}$. In many cases, researchers are interested in phenomena progressing over the time scale of viscous diffusion or even longer time scales, such as those relating to shear properties, electrophoresis, and sedimentation. Therefore, in these researches, the hydrodynamic interactions by sound propagation are often ignored by assuming the incompressibility of fluid where fluid density does not change due to infinite sound speed. However, when we investigate phenomena associated with sound propagation, such as sonic agglomeration, acoustic spectroscopy, and electroacoustic measurements, the hydrodynamic interactions by sound propagation will be important.

In the present research, we investigated the role of sound propagation in the hydrodynamic interactions by considering the compressibility of the fluid. For this purpose, we performed the direct numerical simulations, wherein the hydrodynamic interactions are directly computed by simultaneously solving for the motion of the fluid and the motion of the particle.

2. Discussion: The fluid flow caused by a particle impulsive force was added is shown in the figure below. The velocity field consists of incompressible and compressible components. The strengths of them are represented by the divergence and rotation of velocity, respectively. They are results of different compressibility factors. The rotation remains nearly constant, regardless of the

compressibility, which indicates that the incompressible component is not affected by the compressibility. The effect of compressibility is observed only from the divergence of the velocity field. Each of components gives different flow field patterns. The pattern of the total velocity field is described as the superposition of the vortex convection of the incompressible component and the source-sink flow of the compressible component. The compressibility factor governs the relative time evolution of each component to produce various flow patterns. This difference is reflected in the different roles of viscous diffusion mode and sound propagation mode of hydrodynamic interactions.

In the presentation, we will show the time evolution of hydrodynamic interactions between two particles. The different role of incompressible and compressible components, namely the viscous diffusion and sound propagation in the hydrodynamic interactions will be obvious.

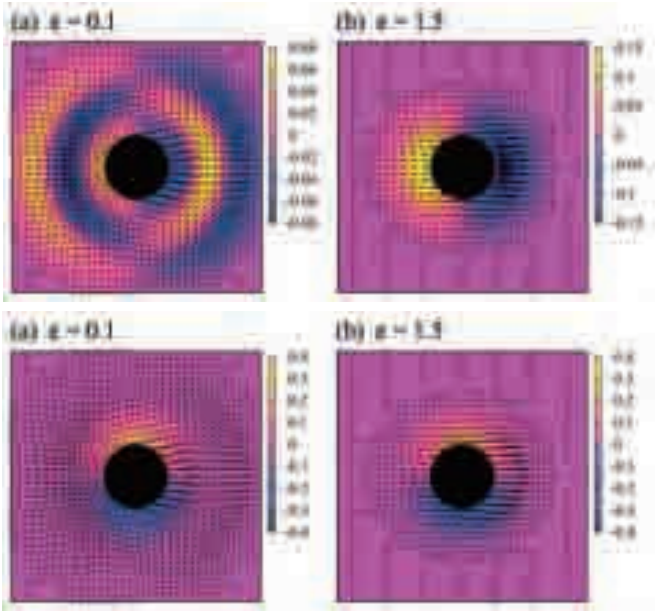


Figure. The flow pattern around a particle. An impulsive force is added on the particle at $t = 0$. The direction of the initial particle velocity is right in these pictures. The time of the figure is $t/\tau_v = 0.27$. The color scale represents the divergence (above: $\nabla\cdot\boldsymbol{v}$) and the rotation (below: $|\nabla\times\boldsymbol{v}|$) of this flow. The divergence is related to the fluid density deviation and inherent for compressible fluids. (It is zero for incompressible fluids). The compressibility factors are (a) $\varepsilon = 0.1$ and (b) $\varepsilon = 1.5$.

Reference: R. Tatsumi and R. Yamamoto: Phys. Rev. E **85**, 066704 (2012).

P-06

Direct Numerical Simulations of Sedimentation Using a Smooth Profile Method

Adnan Hamid and Ryoichi Yamamoto
Department of Chemical Engineering, Kyoto University, Kyoto 615-8510

Summary: Sedimentation of mono-disperse spherical particles at low volume fraction is studied by direct numerical simulations, using a smooth profile method at finite Peclet number (Pe). We investigated the relative effects of thermal and hydrodynamic interactions on velocity fluctuations and diffusion and found a clear transition from a Brownian motion dominant regime to a hydrodynamic dominated regime. We found that both velocity fluctuations and self-diffusion increase with increasing Pe. Furthermore, the diffusion anisotropy is also found to increase with increasing Pe, saturating at higher Pe.

1. Introduction: The settling of particles in a viscous fluid is one of the fundamental phenomena in the non-equilibrium suspension dynamics but is extremely hard to tackle; owing to the difficulty involved in accurately characterizing the long ranged many body interactions between the particles. These inter-particle interactions, caused by fluid flow, are referred to as hydrodynamic interactions (HIs). They may decay as slowly as $1/r$, strongly affecting the dynamic behavior of the suspension. Although many experimental [1][2], theoretical [3] and numerical [5][6][7] investigations have been performed to provide a basic understanding of sedimentation, knowledge of the long-range interactions between particles remains incomplete. The first significant progress in sedimentation research was made by Stokes who calculated the sedimentation velocity of a single sphere as, $V_s^{\phi=0} = 2ga^2(\rho_P - \rho_f)/9\eta$.

Velocity fluctuations and diffusion are two key parameters of sedimentation. A paradox exists between the experimental and theoretical results, where former suggests that velocity fluctuations are independent of system size, whereas latter found that these converge to a finite value depending on the system size. This contradiction was solved by introducing the concept of a characteristic swirl size or correlation length; i.e., velocity fluctuations are explained as a function of cell size, only when the cell size is less than the swirl size [2]. Self-diffusion is another important parameter of the sedimentation owing to its importance in reaction engineering and mixing. Similar to velocity fluctuations, diffusion is also strongly anisotropic, and the vertical diffusion is greater than the horizontal diffusion. This anisotropic behavior arises from the difference in magnitudes of relaxation times of the vertical and the horizontal velocity fluctuations.

The non-equilibrium properties of colloidal particles are difficult to simulate because they involve long-range many bodies HIs that require enormous temporal and spatial scales. A number of different methods are available to simulate the non-equilibrium colloidal phenomena, we performed the direct numerical simulation, using a smooth profile method (SPM) [8][9][10], which replaces the original sharp boundaries between the particle and the host fluid with a diffuse interface of finite thickness. This approach enables us to use a fixed Cartesian grid, thereby significantly improving the numerical computations. In order to properly express the shape of particles, researchers have used different kind of meshes like, unstructured mesh, over-set mesh and boundary fitted mesh. However, the reconstruction of theses meshes at each simulation step and the boundary conditions applied at the surface of the particles to solve the Navier Stokes equations, make them inefficient. In SP method these problems are solved by introducing a smooth profile function of the form, $0 \le \phi(\mathbf{x},t) \le 1$, where $\phi = 0$ stands for fluid domain, $\phi = 1$ indicates the particle domain and $\mathbf{x}$ denotes the special position of the particle. This small change greatly benefits the performance of our simulations as regular coordinates can be used for many particle systems with any particle shape. In addition, the simulation scheme is free of mesh reconstruction at each simulation step. In SPM motion of the colloidal particles is obtained by solving the Newton's equations of motion and solvent motion is obtained by solving the modified Navier-Stokes equation. The direct inter-particle interactions are presented by the truncated Lennard-Jones potential. Present study is focused on the sedimentation at

finite Peclet number, which measures the relative strength of thermal and hydrodynamic fluctuations. Simulations are performed at a volume fraction of 0.02 in a cubic periodic box of size, L/a=32.

2. Results: To differentiate between the effects of hydrodynamic and thermal fluctuations, we calculate the time dependent velocity fluctuations auto correlation functions (VACF). We found that the relaxation times of these correlations are different for the vertical and horizontal directions, whereas the experiments did not show a large difference in these time scales. The difference in time scales arises from the cubic periodic simulations box. This difference in time scales can be reduced by either using the elongated simulation box or increasing the volume fraction. In the experimental study [1], it was found that these correlations relax of the form, $C_\alpha(t) = \left(\Delta V_{\mathrm{H}}^{\alpha}\right)^2 \exp\left(t/\tau^{\alpha}{}_{\mathrm{H}}\right)$, where $\Delta V_{\mathrm{H}}^{\alpha}$ ($\alpha \in x, z$) and $\tau^{\alpha}{}_{\mathrm{H}}$ denote the amplitude and the relaxation time of the hydrodynamic velocity fluctuations, respectively. The values of these two parameters are found by fitting the exponential relaxation form predicted by Nicolai on VACF. Cunha et al. [7] predicted the scaling relations of these two parameters as, $\Delta V_{\mathrm{H}}^{\alpha}/V_s^o = \sqrt{A_1^{\alpha} L\phi/a}$ and $\tau_{\mathrm{H}}^{\alpha}/t_s = A_2^{\alpha}\sqrt{L/\phi a}$, where A_1^{α} and A_2^{α} are the constants that depends on the system configurations. The values of these constant are found from the values of $\Delta V_{\mathrm{H}}^{\alpha}$ and $\tau^{\alpha}{}_{\mathrm{H}}$. We also found that ratio of vertical to horizontal hydrodynamic velocity fluctuations is 2.36, is in good agreement with the experiments [1]. From the asymmetry of the system induced by the gravity, velocity fluctuations are expected to be anisotropic.

Diffusion refers to the fluctuating motion of the particles. To examine the anisotropic particle diffusion in sedimentation, we define the vertical diffusion coefficient as, $D_z(t) = \frac{1}{2t}\left\langle \left(R_{iz}(t) - R_{iz}(0) - V_{sed}t\right)^2 \right\rangle$ as shown in Fig. 1(a). It shows that the long time steady-state diffusion coefficient ($D_z = \lim_{t\to\infty} D_z(t)$) increases with increasing Pe. To understand the diffusivity, we consider the total diffusion coefficient as the sum of thermal and hydrodynamic diffusion coefficient ($D_\alpha = D_o + D_{\mathrm{H}}{}^{\alpha}$), which can be estimated by aforementioned scaling arguments as,

$$D_{\mathrm{H}}^{\alpha} \approx \left(\Delta V_{\mathrm{H}}^{\alpha}\right)^2 \tau_{\mathrm{H}}^{\alpha} = A^{\alpha}{}_1 A^{\alpha}{}_2 V_s^o a\phi^{1/2}(L/a)^{3/2}, \tag{1}$$

$$D_z / D_o = 1 + 4.313Pe, \tag{2}$$

$$D_x / D_o = 1 + 0.158Pe, \tag{3}$$

where D_o is the equilibrium diffusion coefficient which can be obtained via Einstein relation. These scaling relations indicate that the vertical and horizontal diffusion coefficients increase linearly with Pe, but with a smaller pre-factor in the horizontal direction. We can use these scaling relations to predict the diffusion anisotropy as shown in Fig. 1(b). It shows that anisotropy increases with the increase of Pe, saturating at high Pe. We found that the saturated value remain unchanged at low volume fraction. A deviation from the predicted scaling is evident at low Pe, showing the dominant effect of thermal fluctuations. Our simulations show good agreement with the scaling. The anisotropy achieved is relatively higher than the one reported in the experimental study [1]. This higher value is attributed to the cubic periodic box and periodic boundary conditions used.

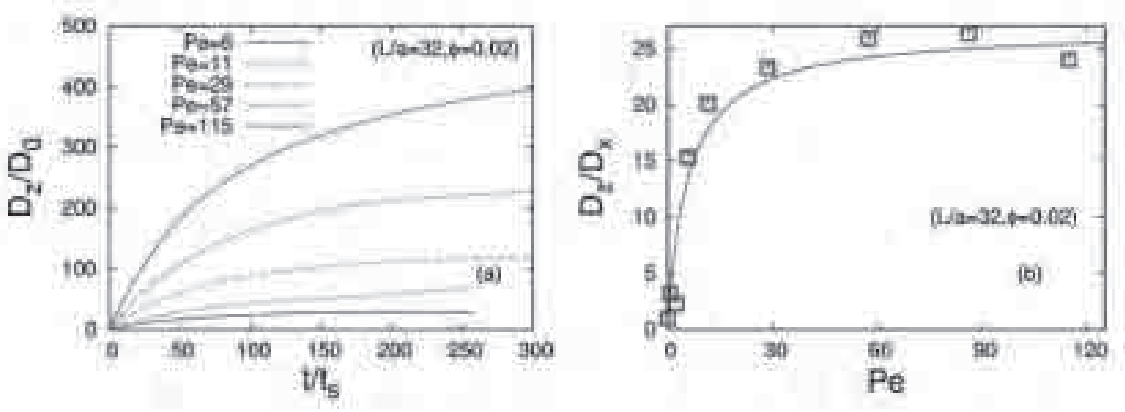


Figure 1: Vertical self-diffusion normalized by the equilibrium diffusion coefficient (a) and anisotropy in vertical and horizontal self-diffusion coefficients (b) at a system size L/a=32 and volume fraction of 0.02 are shown.

3. References:

[1] H. Nicolai, B. Herzhaft, E. J. Hinch, L. Oger and E. Guazzelli, Phys. Fluids **7**, 12 (1995).

[2] P. N. Segre, E. Helbolzheimer and P. M. Chaikin, Phys. Rev. Lett. **79**, 2574 (1997).

[3] R. E. Caflish and J.H.C. Luke, Phys. Fluids **28**, 759 (1985).

[4] D. L. Koch and E. S. G. Shaqfeh, J. Fluid Mech. **224**, 275 (1991).

[5] J. C. Ladd, Phys. Fluids **9**, 491 (1997).

[6] J. T. Padding and A. A. Louis, Phys. Rev. E **77**, 011402 (2008).

[7] F. R. Cunha, G. C. Abade, A. J. Sousa and E. J. Hinch, J. Fluids Eng. **124**, 957 (2002).

[8] Y. Nakayama and R. Yamamoto, Phys. Rev. E **71**, 036707 (2005).

[9] Y. Nakayama, K. Kim and R. Yamamoto, Eur. Phys. J. E **26**, 361 (2008).

[10] T. Iwashita, Y. Nakayama and R. Yamamoto, J. Phys. Soc. Japan **77**, 074007 (2008).

P-07

Molecular Simulations of Polymers under Elongational Deformations

Yuichi Masubuchi and Hiroshi Watanabe
Institute for Chemical Research, Kyoto University

Summary: Although flow property of polymeric liquids such as molten plastic has a significant role in industry, the molecular mechanism has not been elucidated yet. In this study, we focus on the elongational viscosity that is dominant in fiber spinning, film manufacturing, blow molding, etc., since conventional theories cannot explain some experimental data, specifically the difference between melts and solutions; the elongational viscosity monotonically decreases with an increase of flow rate for melts while it increases for solutions beyond a certain critical flow rate. According to the literature data, we proposed a change of molecular friction according to the stretch and orientation of polymer molecule under elongational flow. To check our hypothesis, we determined a functional form of the friction as a function of the stretch/orientation order parameter and examined the determined function in a molecular model to show consistency with the experiments. We also performed molecular simulations for further understanding of the change of friction.

1. Introduction It has been known that flow property of polymeric liquids (such as molten plastic) plays a significant role in industry. For instance, for injection molding (that is widely used to manufacture mass products like mobile phone), less viscous material is preferable to achieve better transferability, higher processing cycle, and reduction of warpage and deformation of the product. However, an adequate high viscosity is also demanded to eliminate burr that is generated by undesirable flows into the gap between molds. The material tuning on elasticity is also important in film blowing, blow molding and foaming where the extrudate needs elasticity to keep its shape after the outlet of extruder.

Since the flow property of polymers is a result of the Brownian motion of molecules, it is strongly dependent on molecular characteristics, i.e., not only chemistry but also molecular weight, its distribution, long chain branching, blending, etc. For instance, when a polymer is in linear shape and its molecular weight is monodispersed, the viscosity is in proportional to M^3 (M: molecular weight). When the material has a distribution in its molecular weight and/or a long chain branching, no simple rule exists to describe the relaxation time as a function of the molecular characteristics.

Because the molecular motion is modified by flow, the flow property also depends on the type of flow and the flow rate. Under shear flow (that is a type of flow generated in the injection molds) the viscosity decreases with an increase of flow rate. On the other hand, under elongational flow (that is dominant in fiber and film processing and blow molding) the flow behavior has not been established yet. For example, monodisperse linear polystyrene melts exhibit monotonic thinning of the steady state elongational viscosity with increasing the strain rate $\dot{\varepsilon}$ even beyond the Rouse relaxation frequency, τ_R^{-1} (this frequency corresponds to the rate of relaxation of the polymer molecule for stretching-contraction motion) [1]. This behavior is quite different from the thinning followed by hardening at $\dot{\varepsilon} > \tau_R^{-1}$ observed for entangled semidilute solutions [2].

2. Friction under elongation We attempt to elucidate the molecular origin of this difference by focusing on the concept of stretch/orientation-dependent monomeric friction ζ recently proposed by Ianniruberto and coworkers [3]. Specifically, literature data of the stress relaxation after cessation of transient elongational flow, reported for both PS melts and solutions, are analyzed to evaluate the stretch/orientation-dependent decrease of ζ.

In our working hypothesis, ζ is expressed as a function of the factor $F_{SO} = \tilde{\lambda}^2 \bar{S}$, where $\tilde{\lambda}$ is the normalized stretch ratio of entangled subchains defined with respect to the fully stretched state, and $\bar{S}$ is an average orientational anisotropy of the components (polymer plus solvent if any) in the system. The factor F_{so} was estimated from the stress decay data after flow cessation as indicated in Figure 1 by solid curve. The resulting functional form of $\zeta(F_{so})$ is described as

$$\frac{\zeta(F_{so})}{\zeta(0)}\frac{1}{f_{FENE}} = \frac{1}{(1+\beta)^{\gamma}}\left\{\beta + \frac{1-\tanh\alpha\left(F'_{so} - F'^{*}_{so}\right)}{2}\right\}^{\gamma}$$

where $F'_{SO} \equiv F_{SO} f_{FENE}$ with $f_{FENE} \equiv 1/(1-\tilde{\lambda}^2)$. α, β, γ and F'^{*}_{SO} are parameters determined by the

fitting as $\alpha = 20, \beta = 5\times10^{-9}, \gamma = 0.15$ and $F'^{*}_{SO} = 0.14$ [4].

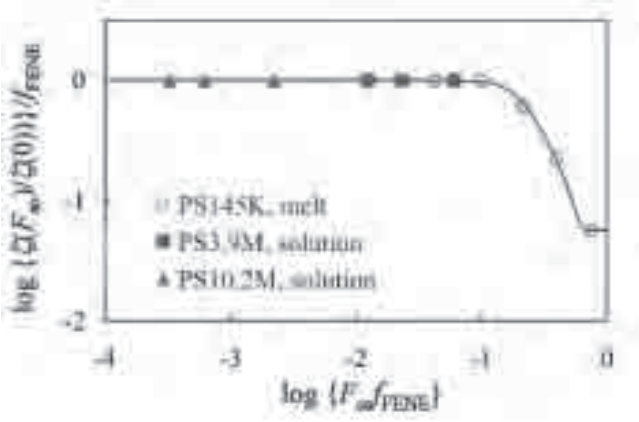


Figure 1 Acceleration factor for PS melt and solutions (from Ref 4).

3. Primitive Chain Network Simulations The $\zeta(F_{so})$ was then used in the primitive chain network (PCN) simulation [5] including finite extensible nonlinear elasticity (FENE) to examine the elongational behavior of melts [4] and solutions [6]. For melts the simulation indicates that ζ decreases significantly under fast elongation because the entangled subchains are short and approach the fully stretched (and fully oriented) limit rather easily. Hence, the steady elongational viscosity η_E follows this decrease of ζ to exhibit the monotonic thinning even at $\dot{\varepsilon} > \tau_R^{-1}$ as shown in Figure 2, left panel. In contrast, for solutions, the simulated η_E exhibits thickening at $\dot{\varepsilon} > \tau_R^{-1}$ (Figure 2, right panel) because the average anisotropy $\overline{S}$ is governed by the solvent and remains small, thus overwhelming the increase of the subchain stretch $\tilde{\lambda}$. The simulated results proved to be in satisfactory agreement with the experiments.

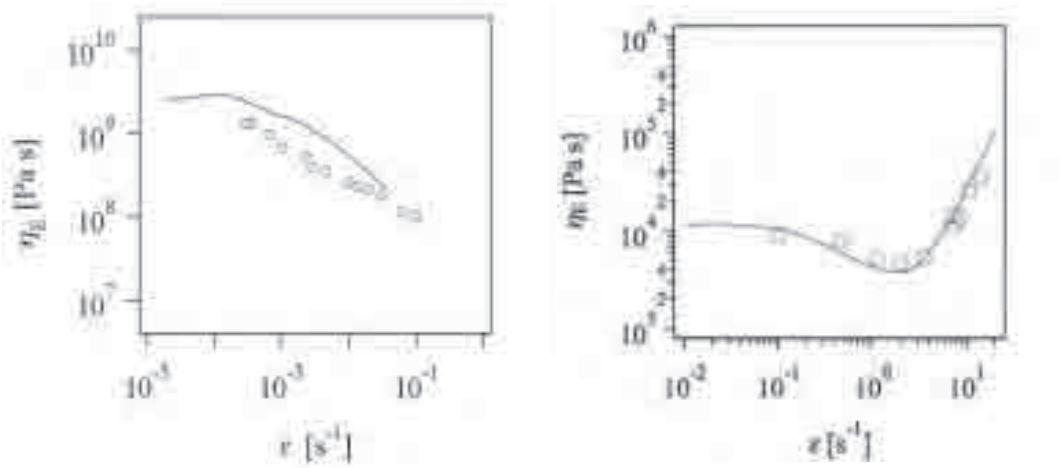


Figure 2 Steady state elongational viscosity of PS melt with molecular weight of 390K at temperature of 130 deg C (left panel) and PS solution with molecular weight of 3.9×10^6 and the concentration of 10wt% (right panel). Symbols are from literature (Ref 1 for melt and 2 for solution) and lines are for PCN simulations with the friction change [4].

The $\zeta(F_{so})$ was also examined for a branch polymer melt [7] and the result for transient elongational viscosity is shown in Figure 3. In the conventional view, the characteristic strain hardening of branch polymers is explained by an excess stretch of the backbone. In addition to this, in comparison to the linear polymer melt, we point that the friction does not decrease much in the branch polymer due to the less orientation of the arms.

4. Kremer-Grest Simulations We further attempt molecular dynamics simulations [8] on the stress relaxation after uniaxial elongations since the change of ζ has been reported for polystyrene melt only (due to the experimental difficulty) so that the effect of chemistry has not been elucidated yet. We used the

standard Kremer-Grest bead spring simulation for a linear polymer melt with the bead number per chain N=40 (No entanglement emerges for this N value). The number density of the bead and the potential parameters (for inter-beads interactions and the spring force between connected beads) are set to follow the standard model. After the sufficient equilibration, uniaxial deformations were applied to the simulation box up to the maximum stretch of 2.0 in Hency strain with various stretch rates. Then the stress relaxation was observed after cessation of the deformation.

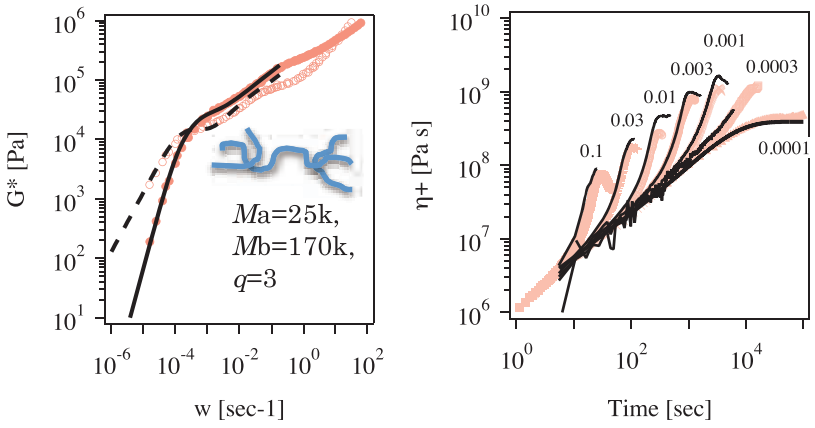


Figure 3 Linear viscoelasticity (left panel) and transient elongational viscosity (right panel) of a pom-pom branched polystyrene melt. Data taken from Ref 7 are shown by symbols and the PCN results with the friction change are shown by curves. Strain rates are shown in the right panel.

Figure 4 shows the stress relaxation after the cessation of deformation. In Figure 4, left panel, the stress is normalized by the value after τ_0 from the flow cessation (at $t = \tau_0$) to eliminate possible contribution from non-bond interaction, which relaxes within τ_0. Clearly, the relaxation is accelerated depending on the stretch rate before the flow cessation. In Figure 4, right panel, the data shown in Fig 1 left are vertically shifted (according to a damping factor h_E) to show that the relaxation rate in the long time range is independent of the deformation rate, and thus, the acceleration occurs only for the higher relaxation modes. These results are consistent with the experiment for polystyrene melt.

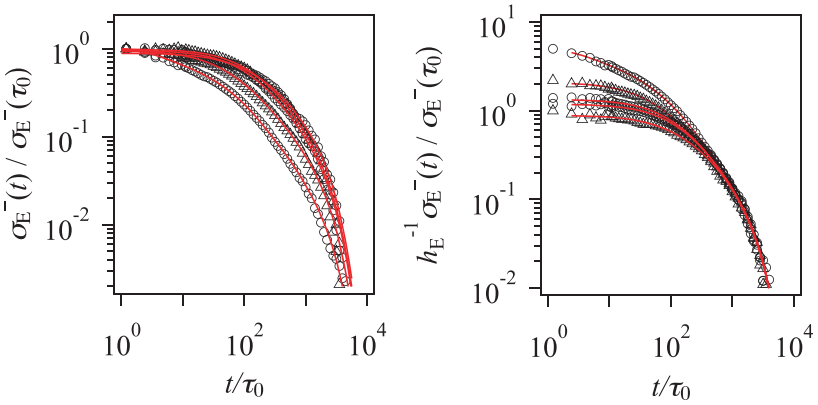


Figure 4 Stress decay after cessation of elongational stretch in the Kremer-Grest simulations (left panel) and test of time-rate factorability at long t (right panel). The stretch rates are 0.05, 0.075, 0.1, 0.2 and 0.5 (σ / τ_0) from top to bottom in left panel and from bottom to top in right panel.

Following our previous study, we performed multi-mode analysis of the stress relaxation data in Figure 4 to extract the acceleration factor. The solid curves in Figure 4 are the fitted multi-exponential functions. The longest relaxation time for the fitting is $1200\tau_0$ that is consistent with the reported value of the longest Rouse time for the Kremer-Grest chain having 40 beads (τ_R= 2 x $1200t_0$= $2400\tau_0$). We applied the procedure X to determine τ_p for higher mode number p. Figure 5 shows the obtained relaxation time τ_p plotted against the mode number p (where p=1 corresponds to the slowest Rouse mode). When the stretch rate is relatively small, the relaxation time obeys the Rouse scaling ($\tau_p \sim p^{-2}$). On the other hand, when the stretch rate becomes higher, fast modes deviate from the Rouse scaling and clearly indicate the acceleration. From τ_p, we calculated the acceleration factor as the ratio of the observed τ_p to the expected value in the

Rouse scaling (shown by the solid curve) at the highest resolved mode, p = 4. The acceleration factor, $\tau_p/\tau_{p,\mathrm{Rouse}}$ involves the effect of FENE in addition to the friction change and is expressed as $\{z(F_{so})/z(0)\}/f_{\mathrm{FENE}}$.

The obtained acceleration factor $\{\zeta(F_{so})/\zeta(0)\}/f_{\mathrm{FENE}}$ plotted against the stretch/orientation factor $F_{so}f_{\mathrm{FENE}}$ is shown in Figure 6. For comparison to the empirical functional form determined from the experimental data (shown by solid curve). It is shown that the Kremer-Grest melt shows less acceleration according to the stress than the experiment. It may be rational because the Kremer-Grest chain has the random force equivalent to each bead. This random force reduces the effective local friction even without the stretch/orientation compared to the real polymers. Nevertheless, further investigation is apparently required for various systems including different chain length, density, interactions, etc to elucidate mechanism of the acceleration.

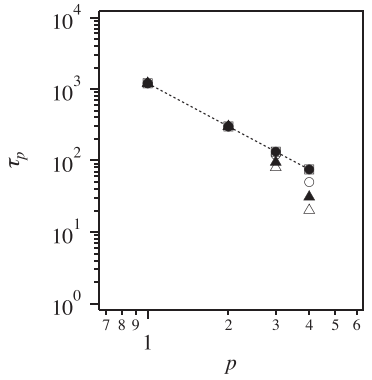


Figure 5 Relaxation times evaluated from the multi-exponential fit of the data in Fig 1. Unfilled and filled triangles, unfilled and filled circles, and square represent the results for the stretch rates at 0.5, 0.2, 0.1, 0.075 and 0.05 (σ/τ_0), respectively. Dotted line indicates the Rouse scaling ($\tau_p \sim p^{-2}$).

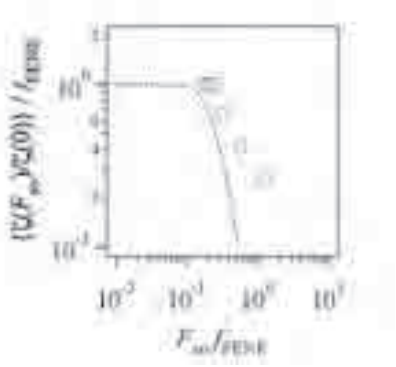


Figure 6 Plot of the acceleration factor $\{\zeta(F_{so})/\zeta(0)\}/f_{\mathrm{FENE}}$ ($= \tau_p/\tau_{p,\mathrm{Rouse}}$) against $F_{so}f_{\mathrm{FENE}}$. Dotted curve is the empirical equation for the polystyrene melt.

References

1. Bach, A.; Almdal, K; Rasmussen, H. K.; Hassager, O. *Macromolecules* **2003**, *36*, 5174.
2. Bhattacharjee, P. K.; Oberhauser, J. P.; McKinley, G. H.; Leal, L. J.; Sridhar, T *Macromolecules* **2002**, *35*, 10131.
3. Ianniruberto, G.; Brasiello, A.; Marrucci, G. *Proc. 7th Annual European Rheology Conf.* **2011**, 61.
4. Yaoita, T.; Isaki, T.; Masubuchi, Y.; Watanabe, H.; Ianniruberto, G.; Marrucci, G. *Macromolecules* **2012**, *45*, 2773.
5. Masubuchi, Y.; Takimoto, J.-I.; Koyama, K.; Ianniruberto, G.; Greco, F.; Marrucci, G. *J. Chem. Phys*. **2001**, *115*, 4387.
6. Yaoita, T.; Isaki, T.; Masubuchi, Y.; Watanabe, H.; Ianniruberto, G.; Marrucci, G. *Macromolecules* **2011**, *44*, 9675.
7. Nielsen, J. K.; Rasmussen, H. K.; Denberg. M.; Almdal, K.; Hassager, O. *Macromolecules* **2006**, *39*, 8844.
8. Masubuchi, Y.; Yaoita, T.; Watanabe, H.; Ianniruberto, G.; Marrucci, G. *Nihon Reoroji Gakkaishi (J. Soc. Rheo. Japan.)* in print.

P-08

Thoughts on Modelling Emergent Phenomena

Sathish K. Sukumaran
Graduate School of Science and Engineering, Yamagata University, Yonezawa, Japan

Summary: After a brief description of the term "emergent", mainly to establish usage, I will describe the main limitation that precludes deducing macroscopic phenomena starting from the microscopic entities by employing brute-force computational approaches and hence, argue for the need for modelling. I will then use examples from fluid flow to illustrate how only a few of the microscopic details and laws are relevant to the modelling of large scale emergent phenomena in the natural sciences. I will then close with the speculation that this is one reason why an understanding of emergent phenomena is highly relevant for the social sciences, where fundamental laws governing the constituent entities are, in any case, hard to come by.

1. Introduction: Many systems of interest**,** either those that fall under the purview of the natural sciences or those from any one of the several mini-universes of human design, consist of a very large number of interacting components. Such systems possess a range of properties and exhibit various behaviours. Certain properties and behaviours of such systems can manifest only in a large collection of such interacting entities and are notably absent when only the individual interacting units are considered. In spite of knowing the characteristics of each of the interacting units and their interactions it is rarely possible to start from this small-scale information and *deduce* the large-scale emergent properties of the system [1-3]. Such properties, behaviours and phenomena can indeed surprise and are usually referred to as emergent [1-5].

Why is the deduction mentioned above rarely successful? As originally envisaged by Laplace [6], one can imagine a situation where armed with the knowledge of how each constituent particle interacts with the others, the starting state of each particle, the equations of motion obeyed by each particle and computers powerful enough to process this information to solve the equations of motion at each instant of time, one would be able to follow the trajectory of each particle in the system over time and thereby simulate any macroscopic phenomena. In fact, what I have provided is a crude description of a technique for simulating physical phenomena on a computer that goes by the name of molecular dynamics. This is wonderful, but only in principle. Although there are several successes by the use of this approach, the method as a whole is rather limited. Even the successes very often rely on inputs other than those mentioned above (I will return to this briefly later). What proves to be the limitation and not easily surmountable is the problem of scale – the size of the system and consequently, the amount of computational memory resources and processing capability necessary to solve problems of

interest. In fact, Laplace himself recognized this. One might argue that computers are becoming bigger (more memory) and faster (increased processing speed). This is true but misses the point. The problem of scale referred to above is truly staggering. In classical mechanics, a description of the state of each particle requires six numbers – three to describe the position of each particle and one each for the momenta along each of the coordinate axes in three dimensional space. Given that the number of constituent particles is of the order of 10^{23} (say, the number of molecules in a few millilitres of water) both the amount of memory required and the number of equations to be solved are astronomically large. This however, is not all. The equations of motion governing *all* macroscopic physical phenomena are believed to be quantum mechanical in nature. If this were so, then the state of a particle can be specified by a certain number of complex numbers, say m. If the system consists of N particles, then we need to specify of the order of m^N complex numbers. In this case, even the specification of the initial state from which to start the simulation becomes prohibitively resource intensive when N becomes even moderately large, let alone anything even remotely close to realistic system sizes [7].

Frequently, it is not necessary to simulate realistic system sizes to investigate the phenomenon of interest. Let us for a moment assume that computational resources are not the limiting factor. Further, even though we do not yet know if starting from a large enough collection of fundamental particles and their interactions, one can actually simulate *all* emergent phenomena [8], for the moment let us assume that this too is possible. Even in this case, the situation would remain unsatisfactory. Although we might be able to simulate any system to obtain the answer to any desired question, we would be as much in the dark as before if the question of interest were to be modified, even if only by a little. If we were to use the simulations just as black boxes, we would have to perform another simulation to obtain the answer to this modified question and so on. This situation obtains because simulations as such do not offer any insight or understanding of the principles governing system behaviour. These higher order principles have to be abstracted from the simulations and incorporated into a condensed representation, one that is capable of both prediction and explanation. In other words, we need models.

Models have proven immensely useful as succinct descriptions of natural phenomena [9]. In addition and more importantly, models endeavour to explain the nature of phenomena – they seek to provide understanding. The typical reductionist approach to understanding, especially in the physical sciences, would be to attempt to “reduce” the phenomenon to laws or principles considered more “fundamental”. In their endeavour to describe and to explain, models also serve as compact records of the hard-won insights of the original researcher into the crux of the phenomenon and thereby provide a rather significant and perhaps, often overlooked additional benefit – they enable the dissemination of these insights to others with relative ease through various communication formats. It is in fact the understanding and the insights that models provide that enable us to answer questions, both those that

are, in some sense, close to the original one but also entirely novel questions, without having to resort to the indignity of performing the whole simulation again for each one of the questions that arise.

Models, almost by definition, do not include all of the details of the original system. They offer "reduced" descriptions, where the significant reduction is in the scale considered and the concomitant level of detail, and are targeted towards a certain class of phenomena. One question that is of significant concern here, especially given the inherent complexity and consequent limitation imposed by scale: why should modelling of emergent phenomena even be possible? What are the features of emergent phenomena that enable the development of such reduced descriptions or models [1-4,10-11]?

The short answer to the above question is that very often, emergent phenomena do not depend very sensitively on all of the multitudinous details of a system to the same extent. In other words, for a range of properties associated with the individual entities that constitute the system and, within reasonable limits, even when the interactions are modified (technically, perturbed), the emergent phenomena are not qualitatively affected. This goes by several names: universality [10.12,13] (especially in the theory of continuous phase transitions [10]), protectorates [2,3,11] etc. Here, perhaps, lies the solution to our predicament and what enables the modelling of emergent phenomena.

In lieu of the above, to model a system, instead of always starting from the fundamental equations of motion of the constituent entities, one usual strategy among researchers is to start with some physically motivated assumptions or principles whose validity and effectiveness are judged by the results they produce. These descriptions are necessarily approximate but still very useful and form the basis not only of models in the physical but also the biological sciences [12]. To probe deeper into this issue, it would be necessary to eschew the general and head towards the concrete. I will do so by considering specific examples from fluid flow.

2. Fluid Flow

Let us start with a specific example, that of fluids. Fluids flow. It is believed that the equations governing fluid flow can, in principle, be derived by starting from a molecular description based on Newton's equation's of motion [14]. However, the equations that describe the spatial and temporal variation of the fluid velocity are found to depend on and can be obtained from a just a few basic ideas. The most important of this, for it is applicable in all cases, is [14,15]:

Some things are conserved: Certain physical quantities cannot be created or destroyed. Hence an increase or decrease of such quantities at any locality in space is accompanied by a corresponding decrease or increase elsewhere. Such quantities include the number of particles and, momenta and energy of the particles.

The conservation laws yield a set of equations that however are not sufficient for developing equations that can describe fluid flow [14]. We need another relationship – the so called, constitutive equation. The simplest of these constitutive equations – Newton's law of viscosity – in conjunction

with the conservation laws mentioned above yields the Navier-Stokes equations that describe the flow of “Newtonian” (they obey Newton's law of viscosity) fluids [14].

On the other hand, say we are interested in simulating fluid-like behaviour, but prefer not to start from the equations of motion. Besides the conservation laws, what ingredients are necessary if we choose not to assume a constitutive equation? It turns out two other principles suffice [15]:

All directions are the same: A fluid has no inherent preferred direction, one along which any of the microscopic processes are different. The choice of direction has to imposed externally, say by flow. Hence, the properties of the fluid at rest are the same in every direction.

Interactions are local: Even though a fluid contains a large number of particles in relative motion, a particle is directly influenced only by the particles in its immediate surroundings. This combined with the conservation law mentioned earlier leads to what are called local conservation laws.

If we introduce local update rules that satisfy the aforementioned three principles, we can simulate fluid-like behaviour and the appropriate constitutive equation (Newton's law of viscosity) emerges as a consequence. Examples include lattice gas cellular automata and lattice Boltzmann approaches [16].

What about fluids that do not obey Newton's law of viscosity (non-Newtonian fluids)? Such fluids are all around us – the kitchen is an excellent place to find several examples – and an active area of research. Is it possible to sketch similar ideas for such fluids? This category includes a diverse range materials exhibiting a significantly wider variety of phenomena and we really do not know whether this is possible, in general. However, there are specific instances where this has already been accomplished and I will focus on one such example below – the case of the so called entangled polymeric liquids [13,17].

In order to study the generic properties of entangled polymeric liquids on a computer, it is not necessary to resort to simulations starting from atomic level details of a specific chemical species. Rather, one can choose a model that combines computational performance with the features believed to be characteristic of such liquids, ones that determine its behaviour [18]:

Liquid-like ordering at the local scale: This is determined by the interactions between the “particles” and the interactions are chosen to enable the relative motion of the particles and hence fluid flow.

Connectivity and Flexibility: Chain flexibility allows the polymer to change its “shape”. Connectivity ensures that the molecule remains a whole even while changing its shape. The ability of the polymer to change its “shape” endows a single polymer chain with elasticity and this is behind the ability of rubber to undergo large, reversible elongations.

Mutual uncrossability of the chain backbones: This feature is necessary to capture the characteristic features of the “entanglements” [13,17,18].

Simulations (for e. g., of the molecular dynamics variety mentioned earlier) that incorporate the above features but otherwise are not exactly careful about atomic level details can still describe several

distinctive features of entangled polymer liquids [19]. Analytical models exist go further and simplify the many chain problem into a single chain problem. They do this by replacing the mutual uncrossability of the chains by a physically motivated ansatz [13,20] but the model needs one additional input that is usually obtained from experimental data. However, recently, once again employing a physically motivated ansatz inspired by Edwards [20], simulations have managed to establish at least a preliminary connection between the chain structure and entanglements, and thereby can provide the additional input necessary for the analytical models mentioned earlier [18].

The ideas mentioned above can be extended to entangled polymeric networks (rubber) by including just one principle, in addition to those mentioned for entangled liquids – that of links (usually chemical) between the chain molecules at a few locations along their length to form a network. These links convert the fluid into a solid whose elasticity originates from the elasticity of a single molecule alluded to earlier and is of a fundamentally different nature to those of crystalline solids, say, steel. Also, just by ignoring the uncrossability constraint, one can also model unentangled polymeric networks. Such simulations exist and have proven rather successful.

3. Speculation: Using examples from fluid flow, I have argued above that in order to describe and understand emergent phenomena, it is often not necessary to resort to atomic level details. One can instead use principles, usually motivated and abstracted from observations and experiments, as a guide to developing a model. Several other examples from the natural sciences exist [1-4,10-15]. Whether this situation generally obtains is anyone's guess. Even if we restrict ourselves only to those phenomena where this might be possible, it still remains a rather useful strategy to seek principles that can guide the development of a model at a scale appropriate to the phenomena of interest. In addition to experiments and observations, the laws obeyed by the microscopic constituents (atoms etc.) also provide useful guidance in seeking such principles in the natural sciences.

Further, I would venture that, this irrelevance of the microscopic details is one of the crucial reasons why understanding the nature of emergent phenomena and how to model them are relevant to the social sciences. We do not as yet know of any “fundamental” laws obeyed by the constituent entities of social systems nor are we sure that we adequately understand all of the myriad ways these entities interact with each other. Even ignoring the limitations imposed by scale, this rules out any kind of constructive approach where the entire system is put together starting from the individual interacting units. In addition, experiments, even when feasible, are rather difficult. However, if we can identify principles, say using observations – data is plentiful these days – coupled with an understanding of the relationships between simple models and possible behaviour, then perhaps there is hope. Such principles may provide the key to developing models of emergent phenomena in the social sciences that, when coupled with “initial conditions” or “history” could perhaps, gradually lead towards understanding.

4. References

[1] P. W. Anderson, Science **177**, 393–396 (1972).

[2] R. B. Laughlin and D. Pines, PNAS **97**, 28–31 (2000),

[3] R. B. Laughlin, Rev. Mod. Phys. **71**, 863–874 (1999).

[4] J. H. Holland, *Emergence: From Chaos To Order* (Basic Books, 1999).

[5] I am aware that my brief discussion of the term "emergence" hardly does justice to the range of prevailing views. I will forego additional discussion but mention that I will use the word emergence in the way it is usually understood in the physical sciences [1,2,3]. This is referred to as weak emergence in the philosophical literature. See for example, D. J. Chalmers, in *The Re-Emergence of Emergence*, edited by P. Clayton and P. Davies (Oxford University Press, 2006) or M. A. Bedau, Noûs **31**, 375–399 (1997). I will not be concerned with strong emergence in this article.

[6] P. S. M. de Laplace, *A Philosophical Essay on Probabilities* (Cosimo Classics, 2007), page 4.

[7] D. M. Ceperley, Rev. Mod. Phys. **71**, S438–S443 (1999).

[8] P.-M. Binder, Nature **459**, 332–334 (2009).

[9] N. Gershenfeld, *The Nature of Mathematical Modeling* (Cambridge University Press, 1998).

[10] N. Goldenfeld, *Lectures On Phase Transitions And The Renormalization Group* (Westview Press, 1992).

[11] R. B. Laughlin, D. Pines, J. Schmalian, B. P. Stojković, and P. Wolynes, PNAS **97**, 32–37 (2000).

[12] R. C. Strohman, Nature Biotechnology **18**, 575–576 (2000).

[13] T. C. B. McLeish, Advances in Physics **51**, 1379–1527 (2002).

[14] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases*, 3rd ed. (Cambridge University Press, 1991).

[15] N. Goldenfeld and L. P. Kadanoff, Science **284**, 87–89 (1999).

[16] B. Dünweg and A. Ladd, in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, edited by C. Holm and K. Kremer (Springer Berlin / Heidelberg, 2009), pp. 89–166.

[17] Polymers are molecules made by connecting several smaller molecules together, typically chemically. When several of these long molecules occupy a small region in space, they cannot move freely because the other molecules get in the way. Such constraints go by the name of entanglements.

[18] R. Everaers, S. K. Sukumaran, G. S. Grest, C. Svaneborg, A. Sivasubramanian, and K. Kremer, Science **303**, 823 –826 (2004).

[19] K. Kremer, S. K. Sukumaran, R. Everaers, and G. S. Grest, Computer Physics Communications **169**, 75–81 (2005).

[20] S. F. Edwards, British Polymer Journal **9**, 140–143 (1977).

P-09

Sparse Multiple Graph Integration for Label Propagation

Masayuki Karasuyama
Bioinformatics Center, Institute for Chemical Research, Kyoto University

Summary: We consider a statistical estimation problem over the graph: Estimating labels of nodes of the graph. Label propagation is a widely accepted approach for this problem that predicts labels of nodes so that they are smooth over the graph. In many recent real-world applications, the same data can be represented by multiple heterogeneous data sources. For example, in gene function prediction, various information sources are available such as gene expression, gene sequences and subcellular localization, which can be all given as graphs. We address the issue of combining multiple graphs under the framework of label propagation. A key challenge under this setting is to integrate different data representations automatically to achieve better predictive performance. The most unique feature of our approach is the sparsity of graph weights which allows to eliminate graphs irrelevant to classification automatically if they are inputted, and further to improve the predictive performance and provide the interpretability of the resultant integrated graphs. We provide an optimization problem formulation, giving weights over input graphs, and an efficient algorithm for solving the problem. We demonstrate the performance advantage and the clear interpretability of our approach through various synthetic and two real-world datasets.

In the area of the statistical learning, graph based semi-supervised learning has received significant attention, due to the flexibility and easiness of implementation, been widely used in many practical applications. A well accepted approach of graph based semi-supervised learning is label propagation (e.g., [1-9]) which propagates labels of each node in a given graph to their neighboring nodes according to their similarity and can be formulated as the simple least squares problem.

The usefulness of the label propagation algorithms has been demonstrated so far, but their performance highly depends on how we generate the input graph. That is, we need to choose the similarity measure (or the distance function) which determines edges of the graph. More importantly, in many practical applications, we have a number of different graph data sources. For example, in gene function prediction, various information sources are available such as gene expression, gene sequences and subcellular localization. We then cannot see the most important graph for prediction. In this paper, we address the issue of integrating multiple graphs under the label propagation framework, where we do not need to choose one specific similarity measure and one data source a priori. Instead, the algorithm automatically estimates the optimal combination of given multiple graphs.

We propose a new approach for integrating multiple graphs under the label propagation framework. As already done by the most existing methods (e.g., [10-13]), our approach also combines multiple graph Laplacian matrices linearly and estimates their weight coefficients. However, our unique property is the sparsity of graph weights. That is, graph weights of our approach can be sparse, meaning that only a part of weights has non-zero values and the rest are equal to exactly 0. This important property provides the following two advantages:

A) In general, eliminating irrelevant or noisy graphs improves classification performance. Conventional approaches however have cases of assigning some weights to graphs irrelevant to classification, by which prediction performance can be deteriorated if irrelevant graphs are

inputted. On the other hand, our sparseness property allows us to eliminate irrelevant graphs completely because their coefficients are set at zero.

B) Sparse weighted coefficients allows us to identify the graphs which are important (or not needed) for classification easily. Furthermore, our formulation can provide a clear interpretation of the mechanism of sparsity, and it also offers a kind of grouping effect given by elastic net [14].

We demonstrate the performance advantage and the clear interpretability of our approach through various synthetic and two real-world datasets.

Reference:

[1] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in Proceedings of the 18th International Conference on Machine Learning, C. E. Brodley and A. P. Danyluk, Eds. Morgan Kaufmann, 2001, pp. 19.26.

[2] M. Szummer and T. Jaakkola, "Partially labeled classification with markov random walks," in Advances in Neural Information Processing Systems 14, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2001, pp. 945.952.

[3] T. Joachims, "Transductive learning via spectral graph partitioning," in Machine Learning, Proceedings of the 20th International Conference, T. Fawcett and N. Mishra, Eds. AAAI Press, 2003, pp. 290.297.

[4] X. Zhu, Z. Ghahramani, and J.D. Lafferty, "Semisupervised learning using gaussian fields and harmonic functions,"Proceedings of the 20th International Conference on Machine Learning, eds. by T. Fawcett and N. Mishra, pp.912-919, AAAI Press, 2003.

[5] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Sch¨olkopf, "Learning with local and global consistency,"Advances in Neural Information Processing Systems 16, eds. by S. Thrun, L. Saul, and B. Scholkopf, pp.321-328, MIT Press, Cambridge, MA, 2004.

[6] M. Herbster, M. Pontil, and L. Wainer, "Online learning over graphs," in Proceedings of the 22nd Annual International Conference on Machine Learning. ACM Press, 2005, pp. 305-312.

[7] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: from transductive to semi-supervised learning," in Machine Learning, Proceedings of the 22nd International Conference, L. D. Raedt and S. Wrobel, Eds. ACM, 2005, pp. 824-831.

[8] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," Journal of Machine Learning Research, vol. 7, pp. 2399-2434, 2006.

[9] Y. Bengio, O. Delalleau, and N. Le Roux, "Label propagation and quadratic criterion," in Semi-Supervised Learning, O. Chapelle, B. Scholkopf, and A. Zien, Eds. MIT Press, 2006, pp. 193-216.

[10] K. Tsuda, H. Shin, and B. Scholkopf, "Fast protein classification with multiple networks," Bioinformatics, vol. 21, no. suppl 2, pp. 59-65, sep 2005.

[11] T. Kato, H. Kashima, and M. Sugiyama, "Robust label propagation on multiple networks," IEEE Transacations on Neural Networks, vol. 20, pp. 35-44, 2009.

[12] M. Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, "Unified video annotation via multigraph learning," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, pp. 733-746, 2009.

[13] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, and Q. Morris, "GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function," Genome Biology, vol. 9, no. Suppl 1, p. S4, 2008.

[14] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of The Royal Statistical Society Series B, vol.67, no.2, pp.301-320, 2005.

P-10

Integer programming method for completing ortholog-based gene-protein-reaction network by gene essentiality

Takeyuki Tamura
Kyoto University

Summary: In this work, metabolic networks and gene-protein-reaction (GPR) networks are integrated and formalized as penta-partite Boolean networks (PPBN) to analyze relations between genes and reactions, and between gene knockouts and gene essentialities. We also formalize an optimization problem to reconcile gene essentialities confirmed by biological experiment and the results of computer simulation on this PPBN by appropriately assigning Boolean functions to GPR part of PPBN. Though this problem is proved to be NP-complete, we develop an integer programming based method for the problem. As a result of computational experiment, accuracy of 97.8% was obtained for the data of central metabolism of *Escherichia coli* K-12 MG1655, where the data of KEGG PATHWAY and KEGG ORTHOLOGY are used for the computational experiment and the gene essentialities of Keio collection are used as the biologically confirmed data.

1. Introduction : Knowledge completion by high-throughput data is becoming more important in bioinformatics. In particular, completion of metabolic networks to obtain better metabolic reconstruction is one of intensely studied topics since much data of metabolic networks is available on some public databases such as KEGG [1], EcoCyc [2], etc. However, since these databases are not yet perfect, there is a case that some compounds are not producible according to the knowledge of databases but confirmed producible by biological experiments. The basic notion of gap-filling is to reconcile such inconsistency by inserting new reactions and/or reversing the direction of reactions [3].
As the next step of the completion of metabolic reconstruction, cell growth rate predicted by flux balance analysis (FBA) and gene essentiality confirmed by biological experiments are often used to be compared. FBA is a widely used approach for studying steady states of metabolic networks. By calculating the flow of metabolites, FBA can be used to predict cell growth rates under various conditions [4]. In this method, the cell growth rate is often modeled by the biomass objective function, which reflects the rate of all biomass precursors calculated by FBA [5]. In other words, amount of all related metabolites in a cell is summed up and used as the approximate size of the cell.

On the other hand, screening of cell growth rates of single-gene knockout mutants of microorganisms is also a powerful tool of obtaining data for completing metabolic networks. For example, the cell growth rate of 3985 single-gene mutant of *Escherichia coli* K-12 under the rich nutrient condition, called LB medium, is available. Since a gene-knockout often causes the inactivation of some reaction in a metabolic network and then affects the flux balance, the cell growth rate of single-gene mutant is often very different from that of the wild-type.

Since it is expected that the results of computer simulation and biological experiment match each other, gene essentiality of single knockout is often used to validate metabolic reconstruction datasets such as iAF1260 and iJO1366 [6,7], which are datasets of *Escherichia coli* K-12 MG1655 tuned for FBA. The accuracy of gene essentiality prediction of iAF1260 is 92% for 1260 genes though their reconstruction process may already include some completion from the view point of gene essentiality.
GrowMatch [8] is a method of reconciling growth prediction based on iAF1260 and biologically confirmed gene essentiality data including [9]. A unique aspect of Model SEED is its capability of inferring the biomass objective function of an organism [10].

The method of [11] suggests corrections to the sets of reactions, biomass metabolites, nutrients and secretions. In the above metabolic reconstruction and completion methods, gene to protein to reaction associations (GPR) [12] are used to determine which reactions are inactivated by the knockout of which genes. Though GPR can be represented by AND/OR Boolean network (BN) in many cases, existing completion methods seem not to focus on details of this BN.

In this work, we formalize GPR as an AND/OR tri-partite BN (GPR-TBN) and then define *GPR-metabolic network* as penta-partite BN, where reactions of a metabolic network is regulated by GPR-TBN. We also formalize a optimization problem, called GPRmetaCMPL, to maximize the accuracy of GPR-metabolic network by appropriately assigning Boolean functions to GPR-TBN, where accuracy means how much percent of results of gene knockout simulation is consistent with gene essentiality confirmed by biological experiment. Though GPRmetaCMPL is proved to be NP-complete, we develop an integer programming (IP) based method for this problem. Computational experiment is also conducted with data of central metabolism of *Escherichia coli* K-12 MG1655 downloaded from KEGG and gene essentiality data of [9]. As for GPR-TBN, data of KEGG ORTHOLOGY is used. As a result of the computational experiment, accuracy of 97.8% was obtained. Our proposed completing method does not insert or delete edges, but assign appropriate Boolean functions to GPR-TBN. Similar to existing completing methods, we estimate the effect of each single knockout by computer simulation on metabolic networks and reconcile with gene essentiality by completion. Another feature of our method is that metabolic networks are also represented by AND/OR BN where each reaction or compound corresponds to “AND” or “OR” node respectively. The BN model of metabolic networks is relatively tolerant for imperfectness of data compared to the FBA model. Though the FBA model is affected by lack of data in both upstream and downstream, the BN model is only affected by that of upstream. Furthermore, since we use gene essentiality instead of cell growth rate for the completion, calculating real numbers of the biomass objective function by FBA may not be necessary.

2. Discussion : In this work, we define GPR-metabolic network by integrating GPR and metabolic network. We also define the accuracy of the GPR-metabolic network from the view point of the gene essentiality. The problem GPRmetaCMPL is to optimize the accuracy of a GPR-metabolic network by appropriately assigning either “AND” or “OR” function to each node of protein nodes and determinant nodes. Since GPRmetaCMPL is proved to be NP-complete, we formalize the problem as integer programming (IP), for which many efficient solvers are available. Since every constraint and the objective function must be represented by linear equality or linear inequality in IP, we develop a method of representing assigning a Boolean function to a node by linear inequalities. In this method, to avoid multiple valid assignments, we introduce the notion of time. We conducted computer experiments with the data of central metabolism of *Escherichia coli* K-12 MG1655. The data of KEGG PATHWAY and KEGG ORTHOLOGY is used for the computer simulation on the GPR-metabolic network whereas the data of Keio collection is used for biologically confirmed gene essentialities. The accuracy of the computer experiment was 97.8% and the computation time was 15.33 seconds for the network with 63 reactions, 59 compounds, 5 essential genes and 84 non-essential genes. Instead of the biomass objective function in FBA, we use a pseudo biomass objective function, in which we choose 7 reactions and treat as cell death if 4 of the 7 reactions become inactive. The obtained high accuracy implies that modeling gene-reaction relation by ortholog data is efficient. To apply this method for genome scale gene essentiality data, reducing the number of

variables in IP is one of our future works since the notion of time drastically increases the number of variables. To handle this problem, decomposing cycles by feedback vertex set may be efficient.

References

[1] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe, M: KEGG for integration and interpretation of large-scale molecular datasets, **Nucleic Acids Research** 2012, **40**:D109-D114.

[2] Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD: EcoCyc: a comprehensive database of Escherichia coli biology, **Nucleic Acids Research** 2011, **39**:D583--590.

[3] Kumar VS, Dasika MS, Maranas CD: Optimization based automated curation of metabolic reconstructions, **BMC Bioinformatics** 2007, **8**:212.

[4] Orth JD, Thiele I, Palsson BO: What is flux balance analysis?, **Nature Biotechnology** 2010, **28(3)**:245--248.

[5] Feist AM, Palsson BO: The biomass objective function, **Current Opinion in Microbiology** 2010, **13**:344--349.

[6] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information, **Molecular Systems Biology** 2007, **3:121**.

[7] Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO: A comprehensive genome-scale reconstruction of Escherichia coli metabolism – 2011, **Molecular Systems Biology** 2011, **7**:535.

[8] Kumar VS, Maranas CD: GrowMatch: An automated method for reconciling in silico/in vivo growth predictions, **PLoS Computational Biology** 2009, **5(3)**:e1000308.

[9] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, **Molecular Systems Biology** 2006, **2**:2006.0008.

[10] Henry CS, DeJongh M, Best, AA, Frybarger PM, Linsay B, Stevens RL: High-throughput generation, optimization and analysis of genome-scale metabolic models, **Nature Biotechnology** 2010, **28**:977--982.

[11] Latendresse M, Krummenacker M, Trupp M, Karp PD: Construction and completion of flux balance models from pathway databases, **Bioinformatics** 2012, **28(3)**:388--396.

[12] Reed JL, Vo TD, Schilling CH, Palsson BO: An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR), **Genome Biology** 2003, **4**:R54.

P-11

Latent Feature Models for Biological Networks

Canh Hao Nguyen
Bioinformatics Center, Institute of Chemical Research Kyoto University

Summary: We study the link prediction problem that given a known network, we derive a statistical model for its structure and use the model to predict new edges (links) of the network. We target the application in biological networks where the links in the networks are partially observed, specifically, protein-protein interaction (PPI) and gene regulatory networks (GRN). We model all these networks with latent features. We propose a method to learn the models efficiently in kernel framework. The method gives very high performances, scales to real data sizes in a little amount of time.

1. Introduction: We study the problem of link prediction, which is to predict whether there is a link between two objects. This is a general problem that can be applied to biological networks such as protein-protein interaction or gene regulatory networks. In these networks, each interaction or regulation is a link we want to predict given the available experimental data for known links. However, to check for all of the links (interactions, regulations) is not feasible. Computational methods would be useful not only to discover new links, but also suggest potential links to speed up biological experiments. Other methods tend to use primitive information such as sequences. Since there is too much redundant information in sequences, computational methods cannot offer a high performance.

2. Materials and Method: We propose to use the supervised learning approach for its scalability to real datasets, given that probabilistic models usually do not scale with data of this size. We formulate the link prediction problem as a problem of classifying pairs of nodes to be in the link class (interaction or regulation) or not. However, usually link prediction methods are not applicable as they are not compatible with the interpretation of these biological networks. Therefore, we explicitly model the network as a graph with latent features. The adjacency matrix of the graph is modeled by products of a latent feature matrix and a feature interaction matrix. The model is named the *latent feature model*. From the model, we build a similarity matrix of nodes in the graph using the ratios of common neighbors. We propose to use Support Vector Machines (SVM) by turning the similarity matrix into a kernel matrix. We build a kernel matrix of pairs of nodes based on the nodes' kernel using pairwise kernels. After that, a SVM is learnt to classify the classes. More details in [2].

We study budding yeast physical PPI network from DIP databases [3]. We also study the other kind of network, GRN of E. Coli from [1]. These are the largest known and reliable networks for each kind, making statistical study more reliable. We study physical interactions as they have the interpretation of protein domain physical binding. Also, for GRN, while they are not known for latent features, we expect that the latent feature model is flexible enough for them as other network models can be expressed in latent feature models as well. To assess the model statistically, we use a reliable part of the networks of the largest connected component of the *m*-core, with different values of *m*.

3. Discussion: We show results of links prediction on these networks as in the Figure 1 and 2.

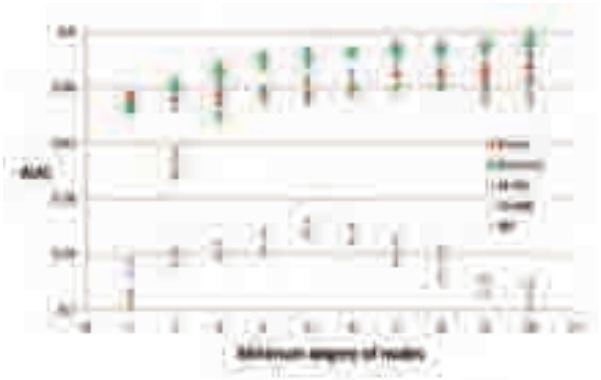


Figure 1: Link prediction results on Yeast's PPI networks.

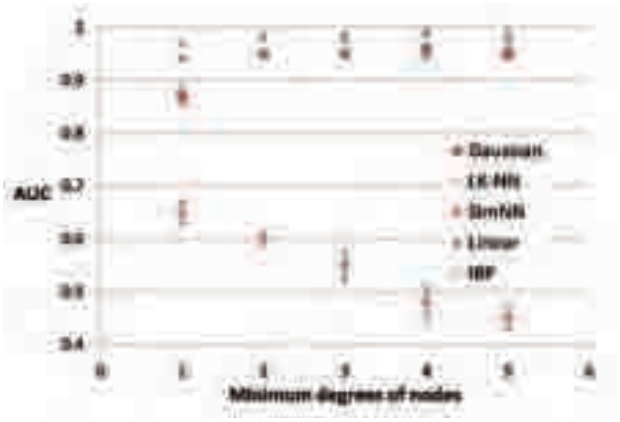


Figure 2: Link prediction results on E.Coli's regultory networks.

We observed the followings. Latent feature models were suitable to model these networks, giving much higher performances than other network models. This also showed that these networks were governed by certain latent features that have not been fully observed and studied. We also observed that our kernel based method was much more efficient than other methods generating latent features explicitly. Our method scaled to real data sizes while the others did not.

References:

[1] Gama-Castro, S. et. al. RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units) *Nucleic Acids Research*. 2010. doi: 10.1093/nar/gkq111.

[2] Nguyen, C.H. and Mamitsuka, H. Kernels for Link Prediction with Latent Feature Models. *Proceedings of ECML PKDD 2011*, LNAI 6912, 2011.

[3] Xenarios,I., Salwinski,L., Duan,X.Q.J., Higney,P., Kim,S.M. and Eisenberg,D. DIP: the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research*, 30, 303305. 2002.

P-12

Identifying Neighborhoods of Coordinated Gene Expression and Metabolite Profiles

Timothy Hancock
Kyoto University

It is reasonable to suggest that changes in gene expression are intended to have a flow on effect on protein and metabolite levels. However, analyzing this relationship is made difficult by competing influences such as post-translational modification and the complex structure of metabolic networks. In this work we investigate how the network structure affects any coordination between transcript and metabolite profiles. To achieve this goal we conduct two complementary analyses focused on the metabolic response of Escherichia coli to stress. First, we investigate the general size of any relationship between networked gene expression and metabolite profiles. Secondly, we employ a novel pathway mining method to investigate the structure of this transcript-metabolite. The results reveal that in general, global regulation signature targeting a small number of metabolites is responsible for a large scale metabolic response.

The dynamics of metabolic networks are the product of complex interactions between genes, proteins and enzymes and metabolites. Since the introduction of DNA microarray technology, the expression signatures of metabolic networks have been extensively analyzed. An underlying assumption of these studies is that fluctuations in gene expression levels are mirrored in the protein and metabolite signals. Although it stands to reason that some relationship exists between metabolic gene expression and other observed metabolic responses, the inherent complexity of metabolism makes the validity of this assumption difficult to assess. Furthermore, it has previously been well established that the correlations between simple gene expression and protein or metabolic flux measurements are unreliable. Recently, many researchers have sought to elucidate these relationships through combined metabolomic and transcriptomic analyses. These combined analyses use techniques such as Gas Chromatography Mass Spectrometry (GC-MS) and microarrays to simultaneously measure changes in metabolite concentrations and gene expression. The integration of these two data sources provides the opportunity to more thoroughly understand how changes in gene expression are converted into metabolic responses.

The results of these combined studies have revealed that transcript and metabolite interaction is often quite complex. Intermediate steps between transcription and metabolite production such as post-

translational modification, regulation or buffering expression by metabolite levels have been found to seriously affect any simple relationship. However, studies have shown simple coordination between metabolite and expression exists although it is either locally restricted, around specific reporter reactions, or highly specific to environmental stress conditions. It is clear that intermediate steps such as post-translational modification and buffering have a pronounced effect on the transcriptome-metabolome relationship. However, the extent to which the network structure of metabolism impacts this relationship is unclear. It is known that metabolic gene expression is highly coordinated along pathways and that this coordinated structure is significantly rewired in response to an external stress. Clearly this regulated coordination of gene expression along metabolic pathways is intended to have a flow on effect within protein and finally metabolite profiles. In this paper we investigate extent to which the network structured coordination of gene expression correlates with metabolite profiles. To address this question we develop models to uncover the gene pathways with the most coordinated expression profiles and then use the expression profiles along these pathways to identify the potential target metabolites.

We evaluate the extent of the transcript-metabolite correlation with a two stage investigation correlating metabolic network gene expression and metabolites for Escherichia coli K-12. In the first stage we tested the hypothesis that transcript-metabolite correlations are sustained at long distances away from the target metabolite. Our results show that overall the transcript-metabolite correlations are sustained over a long network distance of about 3 to 7 genes away from the target metabolites. This suggests that a few hub reactions are controlling the coordinated structure within metabolic networks. We then propose a method to identify these hub reactions by searching for commonly traversed genes over all maximally coordinated expression paths connecting all metabolite pairs. We then use minimum set cover approach to identify hub reactions and metabolites. The result of this analysis revealed that a surprisingly small number of metabolites, approximately 18 to 25, are sufficient to summarize the highly correlated metabolic network structure. Additionally, we found that strong correlations exist between these hub metabolites and reactions and are related to a network regulatory signature.

References:

Hancock T, Wicker N , Takigawa I , Mamitsuka H ,"*Identifying Neighborhoods of Coordinated Gene Expression and Metabolite Profiles.",* (2012) PLoS ONE 7(2): e31345. doi:10.1371/journal.pone.0031345

P-13

Cationic State Distributions over Chlorophyll Pairs in Photosystem I and II

Keisuke SAITO
JST PRESTO/Career-Path Promotion Unit for Young Life Scientists, Kyoto University

Summary: Photosystem I (PSI) and II (PSII) possess chlorophyll pairs P_A/P_B and P_{D1}/P_{D2}, respectively, which are the primary electron donors in the light-induced electron transfer. After the electron transfer, the radical cation remains on these chlorophyll pairs, forming $[P_A/P_B]^{\bullet+}$ and $[P_{D1}/P_{D2}]^{\bullet+}$. The positive charge distributions over the chlorophyll pairs were reported to be asymmetric between the two chlorophylls: ~30/70 for $P_A^{\bullet+}/P_B^{\bullet+}$ and ~80/20 for $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$. We clarify origins of the asymmetric distributions by using theoretical approaches on the basis of protein structures. In PSII, the difference in the electrostatic protein environments between D1 and D2 was significant in determining the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio, whereas geometric differences between P_A and P_B (P_A as the $C13^2$ epimer of chlorophyll *a* and the H-bond pattern) played a role in determining the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio in PSI.

1. Introduction:

In oxygenic photosynthesis of higher plants and algae, protein pigment complexes photosystem I (PSI) and photosystem II (PSII) participate in the conversion of light to chemical energy. Light-induced electron transfer occurs along a series of cofactors bound to the PsaA and PsaB subunits in PSI (Fig. 1a) [1] and the D1 and D2 subunits in PSII (Fig. 1b) [2]. In the center of the protein, there exist chlorophyll pairs P_A/P_B in PSI and P_{D1}/P_{D2} in PSII, which are the primary electron donors in the light-induced electron transfer. Interestingly, P_B, P_{D1} and P_{D2} are chlorophyll *a* (Chl*a*), but P_A is Chl*a*′, the $C13^2$ epimer of Chl*a* (Fig. 2). After the electron transfer, the radical cation remains on these chlorophyll pairs, forming $[P_A/P_B]^{\bullet+}$ and $[P_{D1}/P_{D2}]^{\bullet+}$. The positive charge distributions over the two chlorophylls were reported to be ~30/70 for $P_A^{\bullet+}/P_B^{\bullet+}$ [3-6] and ~80/20 for $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ [5,7,8]

In PSI, there are two electron-transfer branches (A and B), which adopt a pseudo-C2 symmetry with the rotation axis passing through P_A/P_B and F_X. In PSII, not the D2 branch, but the D1 branch serves as

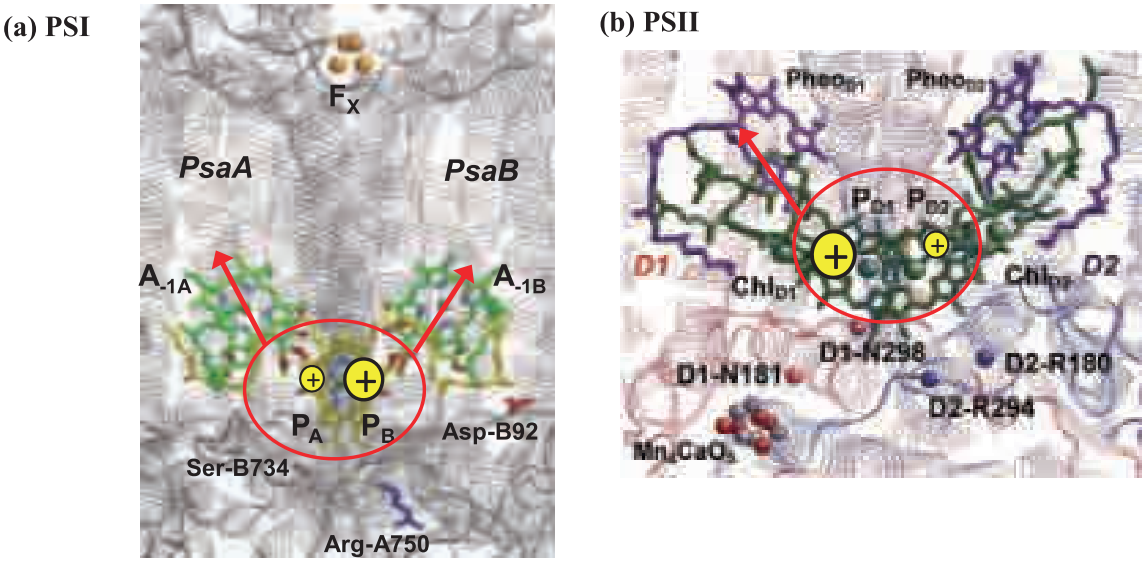


Fig. 1: Structure of (a) PSI with Chl*a*′/Chl*a* pair of P_A/P_B and (b) PSII with Chl*a* pair of P_{D1}/P_{D2}. The arrow indicates the active branch in the light-induced electron transfer.

an electron-transfer active branch; this is in contrast to PSI, in which two branches are electron-transfer active [9,10] (Fig. 1).

So far, a relationship between the activity of the electron-transfer branches and the cationic state population of the Chl*a* pair remains unclear in both the PSI and the PSII systems. An understanding of how the protein environments control the cationic state distribution over the two Chl*a* molecules is prerequisite to elucidate the factors that differentiate the energetics/kinetics between the dual-branch electron transfers in PSI and the single-branch electron transfer in PSII. To understand the role of the amino acid residues or cofactors of PSI and PSII in the energetics of the P_A/P_B and P_{D1}/P_{D2}, we present the following [11,12]: (i) The computational results of the influence of the protein environment on the redox potentials, $E_m(P_A)$, $E_m(P_B)$, $E_m(P_{D1})$ and $E_m(P_{D2})$ are presented on the basis of the crystal structure [1,2] in the presence of all protein subunits and cofactors. The linear Poisson‐Boltzmann equation is solved by considering the protonation states of all titratable sites in the protein. (ii) The $P_A^{\bullet+}/P_B^{\bullet+}$ and the $P_{D2}^{\bullet+}/P_{D2}^{\bullet+}$ ratios for the P_A/P_B and the P_{D2}/P_{D2} pairs, respectively, are calculated using a large-scale quantum chemical/molecular mechanical (QM/MM) approach with explicit treatment of the complete protein atomic coordinates divided into two subsystems; the QM region contains the chlorophyll dimer and is treated by quantum mechanics (unrestricted DFT/B3LYP and LACVP* level) and the remaining protein subunits and cofactors are treated with the MM force field.

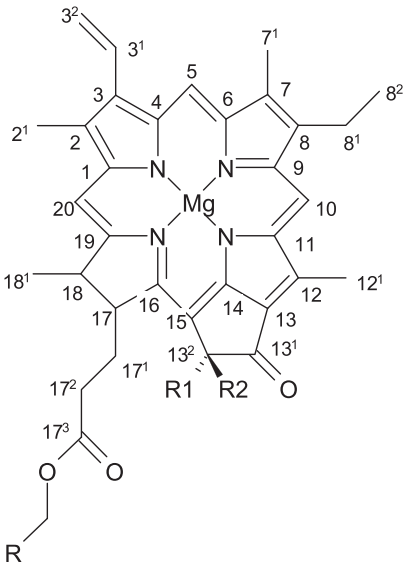


Fig. 2: Geometry of Chl*a* (R1 = CH_3COO-, R2= H) and the $C13^2$ epimer, Chl*a*′ (R1 = H, R2 = CH_3COO-).

2. Results for P_A/P_B in PSI:

$P_A^{\bullet+}/P_B^{\bullet+}$ ratio in PSI. The $P_A^{\bullet+}/P_B^{\bullet+}$ ratio was calculated to be 27.9/72.1 (Table 1a), demonstrating that the cationic state is stabilized more in P_B than in P_A [11]: this should result in the value of $E_m(P_B)$ being lower than that of $E_m(P_A)$. The $P_A^{\bullet+}/P_B^{\bullet+}$ ratio of 27.9/72.1 was considerably close to the charge-distribution ratio of 33/67 obtained from FTIR studies of PSI from *Synechocystis* sp. PCC 6803 [3]. The obtained value was also in good agreement with the experimental spin-distribution ratios of 25/75 [5] and 25/75‐20/80 of PSI from spinach [6] and 15/85 of PSI from T. elongatus [4].

Influence of H-bond on the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio. The -OH group of Thr-A743 can form a H-bond with the 13^1-keto group of P_A, whereas the corresponding H-bond is absent in P_B (Fig. 3a). A mutation of Thr-A743 to Val would result in the loss of the H-bond [13,14]. Using the wild-type PSI crystal structure, we modeled the T(A743)V PSI by substituting the -OH side chain group of Thr with $–CH_3$. $E_m(P_A)$ was lowered by ~10 mV upon the T(A743)V mutation while $E_m(P_B)$ remained unchanged (Table 1a) [11]. As $E_m(P_A)$ was lowered (i.e., $P_A^{\bullet+}$ was stabilized), the calculated $P_A^{\bullet+}/P_B^{\bullet+}$ ratio was significantly shifted and the population of $P_A^{\bullet+}$ increased from 27.9 in the wild-type PSI to 32.2 in the T(A743)V PSI (Table 1a). This observed $P_A^{\bullet+}$ population shift is in agreement with (i) the previous proposal that deletion of the H-bond leads to a downshift in $E_m(P_A)$ [13] and (ii) the experimentally observed relocation of ~14–18 % of the cationic state from P_B to P_A upon T(A743)V mutation in PSI

Table 1: (a) Values of $E_m(P_A)$, $E_m(P_B)$ (in mV), $P_A^{\bullet+}/P_B^{\bullet+}$ ratios in the PSI protein (in %).

	E_m (versus reference)[a]		Charge	
	P_A	P_B	$P_A^{\bullet+}$	$P_B^{\bullet+}$
Wild type	29 [a]	27	27.9	72.1
[H-bond]				
T743V	20 [a]	28	32.2	67.8
[Epimer]				
CH_3COO- deleted			32.2	67.8

[a] Because E_m(Chl*a*′) in the reference model system is unknown, $E_m(P_A)$ was calculated using E_m(Chl*a*) in CH_2Cl_2. An actual E_m(Chl*a*′) may be higher than E_m(Chl*a*) because Chl*a*′ is thermodynamically less stable than Chl*a*.

(b) Values of $P_A^{\bullet+}/P_B^{\bullet+}$ ratios in vacuum (in %).

	Charge	
	$P_A^{\bullet+}$	$P_B^{\bullet+}$
Wild type	41.5	58.5
CH_3COO- deleted	45.5	54.5

from *Chlamydomonas reinhardtii* [14].

Influence of C13² epimer Chla' on $P_A^{\bullet+}/P_B^{\bullet+}$ ratio. The asymmetric charge distribution over P_A/P_B may also result from the asymmetry of the molecular geometry in P_A/P_B. The methyl-ester (CH_3COO-) groups in Chl*a*′ and Chl*a* are oppositely orientated with respect to the chlorin plane (Fig 3a). Except for this methyl-ester orientation difference, the molecular structures of P_A and P_B are essentially identical [1].

To investigate this orientation difference on the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio, we substituted the CH_3COO- groups with H in both P_A and P_B (deletion of CH_3COO-), and calculated the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio. Upon deletion of the CH_3COO- groups the $P_A^{\bullet+}$ population increases by ~5 % relative to the wild-type PSI, resulting in a $P_A^{\bullet+}/P_B^{\bullet+}$ ratio of 32.2/67.8 (Table 1a) [11]. Thus, the orientation difference in Chl*a*′/Chl*a* significantly influences the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio in the PSI protein environment.

Under vacuum conditions, the P_A/P_B heterodimer yielded a $P_A^{\bullet+}/P_B^{\bullet+}$ ratio of 41.5/58.5 (Table 1b) [11], which indicates a considerably delocalized cationic state distribution over P_A/P_B in the absence of the PSI protein environment. The deletion of CH_3COO- from P_A/P_B delocalized the cationic state giving a $P_A^{\bullet+}/P_B^{\bullet+}$ ratio of 45.5/54.5 (Table 1b) [11]; i.e., the cationic state is almost equally distributed over P_A/P_B. It can be concluded that in the absence of the PSI protein environment, the orientation of the CH_3COO- group in P_A is predominantly responsible for the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio.

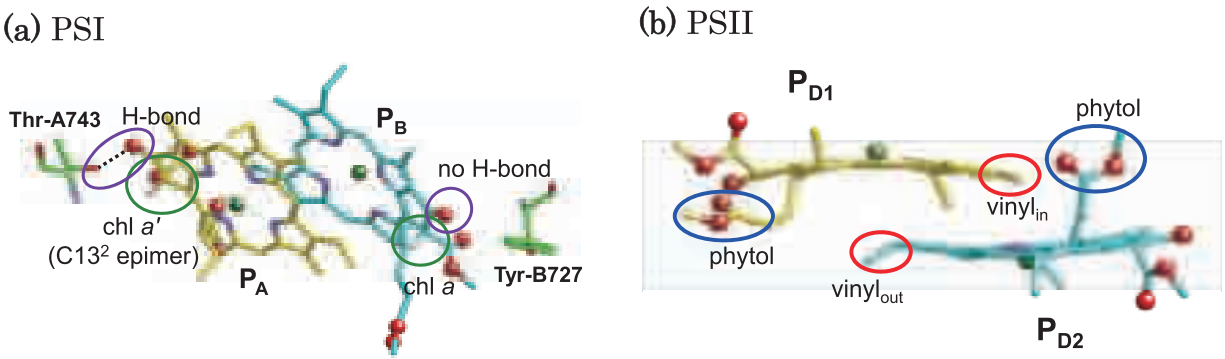


Fig. 3: Geometric difference between the two chlorophylls. **(a)** P_A/P_B in PSI: P_A is hydrogen-bonded to Thr-A743, whereas P_B is not (the purple circle); P_A(Chl *a*′) is the $C13^2$ epimer of P_B(Chl *a*) (the green circle). **(b)** P_{D1}/P_{D2} in PSII: orientations of the phytol group (the blue circle) and the vinyl group (the red circle) are different between P_{D1} and P_{D2}.

Table 2: (a) Values of $E_m(P_{D1})$, $E_m(P_{D2})$ (in mV), $P_A^{\bullet+}/P_B^{\bullet+}$ ratios in the PSII protein (in %).

	E_m		Charge	
	P_{D1}	P_{D2}	$P_{D1}^{\bullet+}$	$P_{D2}^{\bullet+}$
complete PSII	1069 [a]	1167 [a]	76.9	23.1
D1/D2 PSII			71.6	28.4
: Δ(D1-Asn298/D2-Arg294)			61.7	38.2
: Δ(D1-Asn181/D2-Arg180)			64.4	35.6
in vacuum			57.5	42.5

[a] Averaged value of the Mn_4CaO_5 models where the net charge equals 8 [12].

(b) Key residue pairs that increase the $E_m(P_{D1})/E_m(P_{D2})$ difference (more than 20 mV). ΔE_m represents the $E_m(P_{D1})/E_m(P_{D2})$ difference due to the D1/D2 residue pair, i.e., $E_m(P_{D2})$ - $E_m(P_{D1})$.

D1 residues	D2 residues	D1/D2 pair influence		
		$E_m(P_{D1})$	$E_m(P_{D2})$	ΔE_m
D1-Asn181	D2-Arg180	61	110	**49**
D1-Asn298	D2-Arg294	55	98	**43**
D1-Asp61	D2-His61	-31	9	**40**
D1-Glu329	D2-Arg326	-4	32	**36**
D1-Glu189	D2-Phe188	-42	-17	**25**
D1-Asp170	D2-Phe169	-63	-41	**22**

These residue pairs are fully conserved in all known D1/D2 sequences.

3. Results for P_{D1}/P_{D2} in PSII:

$P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio in PSII. The E_m(Chl*a*) values were calculated to be 1069 mV for P_{D1} and 1167 mV for P_{D2} (Table 2a) [12]. The $E_m(P_{D2})$ value was higher than the $E_m(P_{D1})$ value. These results indicate that the positively charged state of the P_{D1}/P_{D2} Chl*a* pair is localized more in P_{D1} than in P_{D2}. Indeed, the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio was calculated to be 76.9/23.1 in the complete PSII [12]. This result is in good agreement with the observed value of ~80/20 for the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio in the previous experimental studies [5,7,8].

The P_{D1}/P_{D2} pair is embedded in the D1/D2 subunits. To investigate the influences of protein subunits and cofactors on the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio, we removed all of the atomic coordinates except for the D1/D2 heterodimer proteins and the cofactors harbored by these two subunits (D1/D2-PSII). In this D1/D2-PSII, the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio was calculated to be 71.6/28.4 (Table 2a) [12], which is not altered significantly compared to the complete PSII. Hence, the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio of ~80/~20 in the entire PSII essentially originates from the D1/D2 heterodimer proteins and the cofactors associated with them. This coincided with the above results that the D1/D2 heterodimer proteins are the major regions that induce the $E_m(P_{D1})/E_m(P_{D2})$ difference due to the differences in the D1/D2 amino acid sequence.

D1/D2 residue pairs that contribute to the larger $P_{D1}^{\bullet+}$ population. The calculated $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio in the D1/D2-PSII indicates that a key to understanding the $E_m(P_{D1})/E_m(P_{D2})$ difference lies predominantly in the D1/D2 environment. The D1 and D2 subunits share high sequence homology to each other. Among the D1/D2 residue pairs, 6 residue pairs contributed to increase the $E_m(P_{D1})/E_m(P_{D2})$ difference by more than 20 mV (Table 2b). Among the residue pairs that contributed to the difference in the $E_m(P_{D1})/E_m(P_{D2})$ values, contributions of the following residue pairs were notable (Table 2b): D1-Asn181/D2-Arg180 and D1-Asn298/D2-Arg294. In fact, vanishing the atomic charges of the D1-Asn298/D2-Arg294 pair resulted in a significant decrease in the occupancy of the $P_{D1}^{\bullet+}$ state, resulting in a $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio of 61.7/38.3 (Table 2a) [12]. Similarly, vanishing the D1-Asn181/D2-Arg180 atomic charges also decreased the $P_{D1}^{\bullet+}$ population.

Influences of geometric asymmetry between P_{D1} and P_{D2} on $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio. If we remove the PSII protein subunits (isolated P_{D1}/P_{D2} pair), the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio (in vacuum) was calculated to be 57.5/42.5 (Table 2a) [12]. In vacuum, the $P_{D1}^{\bullet+}$ population was dramatically decreased relative to that in the PSII proteins but is still the major species. This implies that the P_{D1}/P_{D2} Chl*a* geometries are not identical in the high resolution structure of PSII (Fig. 3b). However, this geometric asymmetry

between P_{D1} and P_{D2} is not so effective on the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio in comparison with an asymmetry between D1 and D2 protein environments.

4. Conclusions:

The $P_A^{\bullet+}/P_B^{\bullet+}$ ratio of the P_A/P_B dimer in PSI and the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio of the P_{D2}/P_{D1} dimer in PSII were calculated to be 28/72 [11] and 77/23 [12], respectively, which are in good agreement with the experimental values. In PSI, the Chl*a*′/Chl*a* difference for P_A/P_B and the H-bond pattern difference (Fig. 3a) played a role in determining the $P_A^{\bullet+}/P_B^{\bullet+}$ ratio [11]. In contrast, the difference in the electrostatic protein environments between D1 and D2, rather than geometric differences between P_{D1} and P_{D2}, was significant in determining the $P_{D1}^{\bullet+}/P_{D2}^{\bullet+}$ ratio in PSII [12]. This difference between PSI and PSII in the origin of the asymmetric distribution of the positive charge on the chlorophyll pairs may be concerned with the factors that differentiate the energetics/kinetics between the dual-branch electron transfers in PSI and the single-branch electron transfer in PSII.

References :

1. Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001) *Nature* **411**, 909.
2. Umena, Y., Kawakami, K., Shen, J.-R., and Kamiya, N. (2011) *Nature* **437**, 55.
3. Breton, J., Nabedryk, E., and Leibl, W. (1999) *Biochemistry* **38**, 11585.
4. Kass, H., Fromme, P., Witt, H. T., and Lubitz, W. (2001) *J. Phys. Chem B.* **105**, 1225.
5. Rigby, S. E. J., Nugent, J. H. A., and O'Malley, P. J. (1994) *Biochemistry* **33**, 10043.
6. Davis, I. H., Heathcote, P., MacLachlan, D. J., and Evance, M. C. W. (1993) *Biochim. Biophys. Acta* **1143**, 183
7. Diner, B. A., Schlodder, E., Nixon, P. J., Coleman, W. J., Rappaport, F., Lavergne, J., Vermaas, W. F. J., and Chisholm, D. A. (2001) *Biochemistry* **40**, 9265.
8. Okubo, T., Tomo, T., Sugiura, M., and Noguchi, T. (2007) *Biochemistry* **46**, 4390.
9. Brettel, K., and Golbeck, J. H. (1995) *Photosynth. Res.* **45**, 183.
10. Joliot, P., and Joliot, A. (1999) *Biochemistry* **38**, 11130.
11. Saito, K. and Ishikita., H. (2011) *Biophys. J.* ***101***, 2018.
12. Saito, K., Ishida, T., Sugiura, M., Kawakami, K., Umena,Y., Kamiya, N., Shen, J. -R. and Ishikita., H. (2011) *J. Am. Chem. Soc.* **133**, 14379.
13. Witt, H., Schlodder, E., Teutloff, C., Niklas, J., Bordignon, E., Carbonera, D., Kohler, S., Labahn, A., and Lubitz, W. (2002) *Biochemistry* **41**, 8557.
14. Li, Y., Lucas, M.-G., Konovalova, T., Abbott, B., MacMillan, F., Petrenko, A., Sivakumar, V., Wang, R., Hastings, G., Gu, F., van Tol, J., Brunel, L.-C., Timkovich, R., Rappaport, F., and Redding, K. (2004) *Biochemistry* **43**, 12634.

P-14

Protein Folding Gone Wrong, Feedback and Parkinson's Disease

Max Brown
Durham University, Biological and Biomedical Sciences

Summary: A third of all cellular proteins are manufactured in one area of the cell called the endoplasmic reticulum (ER), which acts as a protein making factory. Just like a factory, cells have a quality control system involving various feedback mechanisms to get rid of poorly made proteins. Lewy bodies, which are a major hallmark of the neurodegenerative Parkinson's disease, can block this protein quality control system. Research has shown that cells without a functional quality control system can respond by releasing chemicals which cause inflammation. Inflammation is regulated by feedback between cells and the immune system. If feedback which promotes inflammation is maintained the outcome will be long lasting inflammation. This is important because long lasting inflammation in the brain may damage neurons and potentially cause Parkinson's disease.

1. Introduction: Parkinson's disease (PD) is the second most common neurodegenerative disease. It was first described by James Parkinson in 1817. In his monograph entitled *Essay on the shaking palsy* he described the main features of the neurodegenerative condition, which would later be named after him; Parkinson's disease. Nearly two centuries later this disease continues to affect an estimated 1% of people over the age of 65. Not all patients have the same symptoms, but most suffer from rigidity, tremor, postural instability, freezing and slowness of voluntary movement (Dauer and Przedborski 2003). PD is characterised by a loss of dopamine producing cells (known as dopaminergic neurons), which account for the symptoms of the disease. PD is an age-related disease mainly affecting people over the age of 60. As life expectancy in many countries has increased, such diseases have become a greater concern, prompting increased research interests into the causes of PD. However, there is still no known cure and the only current treatments only partially alleviate the symptoms and do not slow or halt the progression of this disease. Most PD cases are sporadic (occurring in a seemingly random manner) and the cause of PD is still not known. However, new emerging evidence has pointed toward neuroinflammation induced by protein folding stress in dopamine producing cells as a cause for PD.

But what is protein folding stress, and how might it lead to PD? Firstly, we must address what protein folding is. Cells have evolved compartmentalisation to allow very specific and complex functions to be carried out in certain parts of the cell. Part of the cell- called the endoplasmic reticulum (ER)- has the role of folding and maturing proteins. This is a very important job as proteins are involved in nearly every cellular process, as well as having a structural role. The ER takes in linear protein strands and folds them to form a functioning mature protein. The ER can be thought of as a cells protein making factory.

However, mistakes can be made and some proteins may not be folded correctly. Mis-folded proteins

are harmful and inhibit the protein production inside the factory (ER) causing it to be in a state of protein folding stress. Fortunately, mis-folded proteins are detected by three different quality control mechanisms within the cells protein making factory (Schröder 2006). These three quality control mechanisms detect the stress and initiate several cellular responses in order to alleviate protein folding stress. Between them they can:

- Initiate the expansion of the protein factory to dilute out the mis-folded proteins (and therefore the damage they can do)
- Increase the amount of protein folding machinery
- Reduce the amount of protein strands entering the factory to reduce the protein folding burden
- Increase protein degradation to remove harmful mis-folded proteins. This protein degradation also recycles proteins back into their original building blocks so they can be used again to make new proteins.

These processes are maintained until homeostasis is restored so that the cells protein factory can go back to 'business as usual'. Once mis-folded proteins are removed, a negative feedback loop turns the protein factories quality control mechanisms off (figure 1, step 1). These quality control mechanisms have been shown to be activated in various models of Parkinson's disease (PD) and in human PD post mortem studies (Smith et al. 2005).

A major pathological feature of PD is the formation of large aggregates, called Lewy bodies. These form in dopamine producing neurons, followed by the subsequent loss of these neurons. These large aggregates have been shown to disturb protein folding homeostasis in three different ways:

- They can block the degradation of mis-folded proteins
- They can inhibit transport of folded proteins out of the protein factory
- They can enter the protein factory and disrupt protein folding machinery

All three will cause an accumulation of proteins in the protein factory. Therefore, the negative feedback which turns off the quality control system will not be activated and the protein factory will remain in a prolonged state of protein-folding stress (figure 1, step 2). This is not good for a neuron, or any type of cell, as protein folding stress can lead to inflammation and cell death. Protein folding stress in neurons may lead to inflammation due to activation of cells of the immune system, known as, microglia, (figure 1, step 3).

There is substantial evidence for neuroinflammation (inflammation specific to the nervous system) in

PD. Inflammation has been described as a ‘double-edged sword’ (Wyss-Coray and Mucke 2002) and neuroinflammation is no exception. Inflammation is a reaction of multicellular organisms, which functions to protect against a range of harmful stimuli including; viruses, bacteria, physical damage and harmful chemicals. Neurons signal their status to microglia through a complex system involving feedback between both cell types. Damaged or unhealthy cells alter their feedback to microglia which results in activation of inflammation (Wyss-Coray and Mucke 2002).

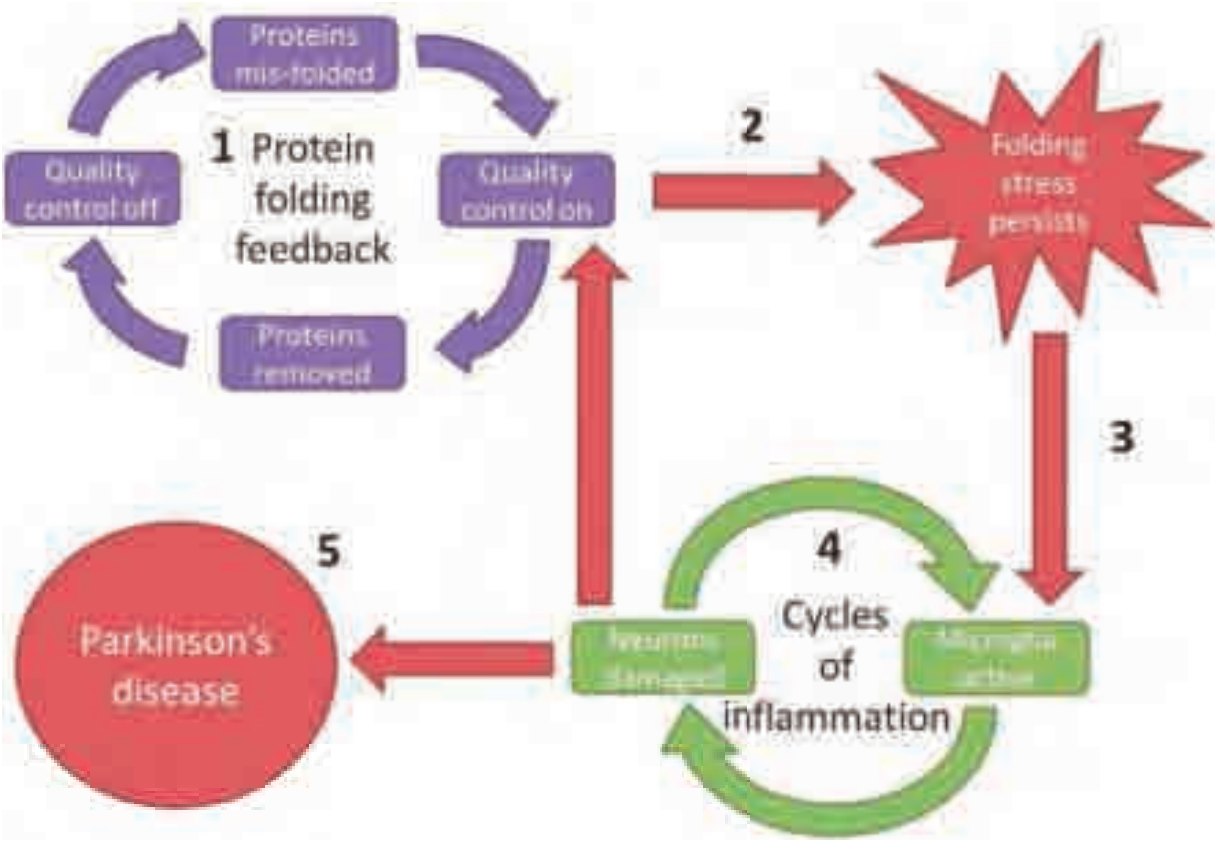


Figure 1. Feedback cycles implicated in PD. Proteins are continually folded in the ER. Mis-folded protein are removed or folded when quality control mechanisms are activated. Once proteins are removed quality control mechanisms are turned off. This protein folding feedback loop maintains normal protein production (1). Lewy bodies disrupt quality control feedback causing persistent protein folding stress (2). Protein folding stress leads to activation of microglia via release of inflammation causing chemicals (3). Activated microglia damage neurons leading to further microglia activation and long lasting cycle of inflammation (4). PD develops as more neurons become damaged and eventually die (5).

In short-lasting inflammation, inflammatory mechanisms promote healing and limit injury. However, prolonged neuroinflammation is detrimental as previously healthy neurons can become damaged by microglia. These damaged neurons will then feedback to the microglia to let them know they are

unhealthy, which could lead to further activation of microglia resulting in a vicious cycle of inflammation (figure 1, step 4). Damage to neurons from microglia may even cause further protein folding stress as microglia can release chemicals, which cause proteins to misfold. This cycle of inflammation and neuronal damage could lead to PD (figure 1, step 5). Prolonged inflammation has been implicated as a cause for diseases such as diabetes and Alzheimer's (Wyss-Coray and Mucke 2002). In the case of PD, inflammation is believed to be initiated by dopaminergic neurons with the initial inflammatory trigger or triggers remaining unclear. Our research goal is to investigate if disruption of natural feedback mechanisms involved in maintaining protein folding homeostasis can contribute to the progression of Parkinson's disease.

2. Discussion: In this biological system feedback has a key role. In one stage, during protein folding (figure 1, step 1), feedback is essential and problems only arise when this feedback is disrupted. In another stage, inflammation (figure 1, step 4), problems occur when the feedback message is altered and is then maintained for long periods of time. The process of maintaining protein production is dynamic and complex involving co-operation between numerous signalling pathways and different parts of the cell. Regulating the immune response to unhealthy neurons is just as dynamic and involves co-operation between different parts of the cell as well as co-operation between cells.

Importantly, this is just one new and emerging hypothesis on the cause of PD. It is likely that these key events may not be the only contributing factors to the progression of PD. However, disruption to protein folding and inflammation feedback mechanisms has been implicated in other neurodegenerative diseases, including Alzheimer's disease, making them important cellular processes worthy of research. Therefore, the identification and development of drugs, which stop alterations to either of the two feedback mechanisms discussed may lead to treatments, which either slow or halt the progression of PD and other neurogenerative diseases.

Reference : This project is funded by Parkinson's UK.

Dauer, W. and Przedborski, S. (2003) 'Parkinson's Disease: Review Mechanisms and Models', Neuron, 29, pp.889-909.

Schröder, M. (2006) 'The unfolded protein response', Molecular Biotechnology, 34, pp. 279-290.

Smith, W. W., Jiang, H., Pei, Z., Tanaka, Y., Morita, H., Sawa, A., Dawson, V. L., Dawson, T. M. and Ross, C. (2005) 'Endoplasmic reticulum stress and mitochondrial cell death pathways mediate A53T mutant alpha-synuclein-induced toxicity', Human Molecular Genetics, 14, pp. 3801-3811.

Wyss-Coray, T. and Mucke, L. (2002) 'Inflammation in neurodegenerative disease- a doubled edged sword', Neuron, 35, pp. 419-432.

P-15

The subjectivity involved in objectively reporting themes and topics in the media

Laurens de Rooij
Durham University

Summary:

The British media coverage of themes has a wide range of perspectives. Developing a greater understanding of their production will aid in the reflection of the position of the Qur'an in the British media. Religious concepts present in the people making media reports i.e. 'reporting agent' influence the report produced. As subjects in contemporary society we are under the influence of a variety of stimulants at any given time. While the media acts as a stimulant to the person receiving it, the media is also influenced by the person producing the media report.

Having accepted the theses of contemporary critical theory and the developments in philosophy, anthropology and psychology in analysing the effect of and on the media. We can argue that it does little to explain the contemporary approach and interaction with the themes reported, by those producing the British media reports. Particularly as one can distinguish religious, socio-economic and ethno-cultural motives which invariably intertwine. This paper will discuss how a reporting agent's perceptions or biases may affect the way a story is reported, as it applies to contemporary reports on various themes in the British media.

1. Long abstract:

In contemporary reporting the receiving subject is influenced by the report and how it treats certain concepts. Contemporary mass media have the ability to reach a wide audience simultaneously with great ease. Not only has the contemporary context reduced the time and space for exchanging information it is also increasingly democratic as certain forms (YouTube, social-networking and blogging) have empowered the individual to becoming a media producing entity. With this all in mind it is important to understand that these processes all affect the coverage of topics in the media. As S. Hall notes, that the media are the sources of both 'primary' and 'secondary' definitions and 'both have a bearing on how the problem will be seen and understood by the public'[1]

Experiments conducted by Elizabeth F. Loftus in the mid to late seventies on the reconstruction of car accidents, led her to conclude that there is direct link between the way a witness

[1] S. Hall, "The Treatment of Football Hooliganism in the Press," in *Football Hooliganism: The Wider Context*, ed. R. Ingham (London: Inter-Action Imprint, 1978), 16.

is questioned and the response he or she gave.[2] In this case it was geared at certain types of verbs describing the collision and at what speed the respondents thought the collision took place. This suggests that when an event is reported by the media a similar transfer of values can take place. This can be conscious, in order to convince the audience of a specific point of view; it can be subconscious and part of the identity of the reporting agent(s). Therefore, if we are to delve deeper into the effect of the coverage by the British media, it is important to be able to trace back the effect of those involved.

However the information is also affected by the interpretation made by the mediator (even if it is unintentional and ever so slightly).This is why two different people could have two different accounts of the same event. Scholars have argued that the production of information to share or broadcast is affected by a variety of intentions before it is disseminated.

There is a need for conceptualisation on how information, be it knowledge gained through direct contact or education as well as pre-conceptions, pre-judices, pre-suppositions and potential misconceptions affect media coverage? The fact that in a globalising world, with modern media outlets, social media and 24/7 news networks, people are bombarded with information in a plurality of ways, affecting their ideas and beliefs in an equally diverse manner.

The media reports concerning a specific issue, receive a variety of reactions by the public and also the press. It is occasionally ignored that the press is often the primary audience that makes up its mind on an issue before relaying that information to the wider public. Nick Couldry states that *Media institutions claim to speak for, or to be our privileged access to 'the social'*.[3] The idea that what is conveyed in the media is not a portrait of the social reality but actually that reality itself. However that reality that is being portrayed in the media is coloured by the lens that is the media. All media reports are coloured by the reporting agent. This can be one reporter on location via live-feed, but it can also be a piece edited by a multitude of people before it is broadcast months after the fact.

The place topics take in the media is a decision made by someone somewhere; it may be a conscious or subconscious decision, but it is a choice. As a result it is mediated, the viewer is being offered a glimpse at reality via a mediated form; it is not reality itself. That is why, for example, seeing a natural wonder on television doesn't really do it justice. One can watch all the documentaries collect all the news footage and cut out and read all the articles on the Angkor Wat but it is still an entirely different experience to when you visit it.

In this public market place of media productions one can refer to their interaction as frames of meaning, interacting, developing and conflicting. The way the media affects the process is not simply

---

[2] Elizabeth F. Loftus, "Leading Questions and the Eyewitness Report," *Cognitive Psychology* 7(1975): 560-72.

[3] Nick Couldry, "Media Rituals: From Durkheim on Religion to Jade Goody " in *Exploring Religion and the Sacred in a Media Age*, ed. C. Deacy and E. Arweck (Ashgate, 2009), 46.

in its ability to broadcast the frames of meaning to large audiences simultaneously but also by shaping the frames of meaning of their audience. However, this state of affairs is still affected by the frames of meaning by the primary individual that reports or broadcasts the event. Important however is that human beings are on the continual quest for meaning and construct it in a variety of ways. Media is simply another domain in which the construction of meaning takes form. It is also arguably the largest source of spreading frames of meaning.

Thus if we think about the spreading of meaning, and with a majority of the media people are exposed to coming from the USA (about 65%), one can ask if the meaning being spread is simply covert cultural imperialism? I feel it is too simplistic to state that, as one does have some alternatives, they might not be many in number or as easily accessible, but outside the USA there is always local broadcasting and broadcasting from surrounding nations. However if you really wanted to and felt strongly enough about it you would not have to watch or read it, you could find the alternative or not watch or read it at all. Yet it is until we all decide we want something different that it will change. The reason it is the way it is, is because the majority, want it that way, accept it or tolerate it. If enough felt it was unbearable then steps would be undertaken to facilitate change, however that in no way should be a reason for violence.

As subjects in contemporary society we are always interacting with the rest of society and increasingly a globalised world. The media influences the person receiving it, but no matter how it is coloured, it affects the rest of society by its impact on people and the effect it has on public debate. Nothing is created *ex nihilo* and as a result there is a need to look beyond the media created and also take into account (if possible) the creator and the process of creation. The manner by which this is done can vary but the primary ideal is that of gaining a greater understanding for the media item at hand situating it in the time in which it was created. Adding to this the analysis of the structures of understanding will allow for an analysis of the interpretation that underlies the creation of a media report. With certain structures guiding interpretation, it is important to gain insight into how the structures of understanding affect certain interpretations. Interpretation is a never- ending requirement of life, the act of interpreting is something humans will do for all time in order to construct meaning into their existence.

In Charles Taylor's work, meaning and interpretation are seen as a continuation of language. Seeing meaning as a constitutive element of language and interpretation is a necessary result of that meaning disclosed within language. This is similar to the effect noticed in the experiments described by Loftus that I referred to earlier. One cannot escape the judgments and processes that are built-into the language one speaks, whether it is a natural language (e.g., English) or a specialized one (e.g.

anthropology, sociology, philosophy etc.).[4] There are intersubjective meanings embedded in the social reality. And these are in turn expressed by individuals in language.[5]

In media reports the Self of the producer crucially affects the way in which the media report is being presented. Therefore even a media report about a given event is not without its subjective influences. This means that media is never received unbiased or unmediated. As S. Hall points out:

> *"Press reports cannot be simply a straight reflection of what happened because there always intervenes a whole process of selection—which events to report, which to leave out; which aspects of an event to report, which to omit; and a whole process of presentation—choosing which sort of headline, language, imagery, photograph, typography to use in translating what happened."*[6]

The media is a very powerful entity in the public domain of modern society. Its influence on the public opinion is considerable, especially when we consider controversial or topics of a sensational nature such as football hooliganism, celebrities, Islam, Muslims and terrorism to name but a few.[7] According to W. Shadid *"The media adds both in a direct and indirect manner to the dissemination of negative imagery concerning allochthonous people and might even play a role in their discrimination by society."*[8] One can see that global media contain valuable assets and many positives. However there are also many dangers and negatives associated with media. There are even issues that can be negative and positive almost simultaneously. It has a vast potential with social networking sites, YouTube etc. that democratise media and bring the world closer together, allowing for people to connect with one another in almost a direct fashion although they are miles away and may never ever meet face to face. This bringing of the world together has created a truly globalised world.

With the media as sources for the spreading of information (and disinformation) it can be the source of producing and enforcing stereotypes and misrepresentation. This is often a result of the under representation of specialised people in the sector and the short-sighted presentation of the issues at hand. [9] As described earlier, to the presentation of the media report is affected by the people

---

[4] Charles Taylor, "Interpretation and the Sciences of Man," *Review of Metaphysics* 25, no. 1 (1971): 3-5.

[5] Chris Mantzavinos, *Naturalistic Hermeneutics* (Cambridge University Press, 2005). 80.

[6] Hall, "The Treatment of Football Hooliganism in the Press," 34.

[7] W. Shadid, "Berichtgeving over Moslims en de Islam in de Westerse Media: Beeldvorming, Oorzaken en Alternatieve Strategieën," *Tijdschrift voor Communicatiewetenschap* 33, no. 4 (2005): 330.

[8] Ibid.

[9] Ibid.

creating them. Thus if they are ill-equipped to deal with the issues at hand (for whatever reason that may be) then the media report is going to reflect that. However at the same time the various media-producing institutions are in a position of power where the effects of their productions can have far-reaching consequences, especially if it is a structural problem rather than a one-off oversight. Henrik Bødker references to a work by Elmelund-Præstekjær and Wien[10] publishing a list of four main necessary criteria, based on case-studies, for a story to be deemed a media event:

1. *The event must be appropriate for public debate, i.e. there must exist a range of legitimate positions as well as people willing to air and debate these.*
2. *The issues at stake must be, and this is linked to the first point, something that can be interpreted within a number of contexts or frames.*
3. *The event must also, at least in their study of a limited number of events, contain some deviation from, a break of, norms.*
4. *The event must be able to condense a complex problem into a striking image and/or draw upon a number of existing stereotypes.*[11]

Poulton, Denham and Lang and Lang use a six stage analysis for the 'agenda building process' that they argue is common in media circles. The stages described are:

1. *The media highlights certain events and issues.*
2. *Different kinds of issues require different levels of news coverage.*
3. *Issues may be framed so as to be understood by news consumers.*
4. *The language used to describe an event or issue may affect public perception.*
5. *The media link the issues and events that have gained salience to 'secondary symbols' that nearly everyone can appreciate or identify.*
6. *Agenda-building accelerates when well-known individuals speak out.*[12]

---

[10] C. Elmelund-Præstekjær and C Wien, "Mediestormens politiske indflydelse og anatomi' [The Political Influence and Anatomy of the Media Storm]," [The Political Influence and Anatomy of the Media Storm.] *Nordicom Information* (2008).

[11] Henrik Bødker, "Muslims in Print, or Media Events as Nodes of Cultural Conflict," in *Media, Religion and Conflict*, ed. Lee Marsden and Heather Savigny (Ashgate Publishing, 2009), 83.

[12] Emma Poulton, "English Media Representation of Football-related Disorder: 'Brutal, Short-hand and Simplifying'?," *Sport in Society: Cultures, Commerce, Media, Politics* 8, no. 1 (2005)., B.E. Denham, "Building the Agenda and Adjusting the Frame: How the Dramatic Revelations of Lyle Alzado Impacted Mainstream Press

The above listed criteria highlight the need for the creators of media to judge the value of the information/experience based on these criteria in order to decide whether it is newsworthy or not. These judgments are taken into consideration even when deciding something as simple as choosing to run a certain story and not another, or the angle at which to approach it from. The media report must not only adhere to certain values instilled in the reporting agent, but also elicit a certain response or set of responses from the prospective audience, which may be positive or negative. Although positive news stories are very rarely mentioned in the press, they do exist, but always with cause for sensation. Everybody loves a hero or an underdog that has triumphed against the odds. Negative events also shape consciousness and elicit negative responses; yet there is no just reason for the media to use biased reporting to promote reasons for ostracising, fearing and in some cases hating groups of people.

Earlier it was described how the media gives insight to the 'real' world and society. By positioning itself in such a way that it offers us unmediated access to reality, the media affects and defines how the individual approaches his or her immediate environment. Thus the media is not only aware of the power it possesses but also uses it to its own advantage.

However it is not only the local news that is reported in local contexts that is affecting local people. In the age of global information technology the news that is reported in one location is also reported around the globe (provided it adheres to the above listed criteria; the events of the past week(s) are a testimony to that.[13] Whether a media report is perceived as positive or negative may also change when it is reported in another location. The same event which may be reported positively in one in one location is reported to show a negative situation in another location or vice-versa.[14] The perception an individual has of the item described in a report is also influenced by the coverage it has received. With Armstrong's arguing that, "*the media have their own agenda, and by their targeting of issues can "make" a problem. The media thus provide the news that fulfils the expectations they create, whether or not they actually elicit fulfilling behaviour. This demonstrates how news values 'work' and*

Coverage of Anabolic Steroid Use," *Sociology of Sport Journal* 16, no. 1 (1999). And G.E. Lang and K Lang, *The Battle for Public Opinion: The President, the Press and the Polls During Watergate* (New York: Columbia University Press, 1983).

[13] For example the publication of certain cartoons depicting the prophet Mohammed or the production of a film for release on internet about the life of the prophet and the events in the history of Islam.

[14] Katherine Pratt Ewing, "Living Islam in the Diaspora: Between Turkey and Germany," *South Atlantic Quaterly* 102, no. 2/3, Spring/Summer (2003): 405-06. and Shadid and van Koningsveld, "The Negative Image of Islam and Muslims in the West: Causes and Solutions," 188-90.

how the media actively engage in news-making".[15]

The media also works toward shaping or creating an audience. This creating of an audience is more than cornering a part of the market by the product on offer. It is offering a type of product and tailoring the audience to consume that product in such a manner that the structures of meaning and understanding are those in relation to and offered up by the object of consumption. The subject's perception of reality is coloured by the media report in such a way that reality cannot be thought of without taking into account the values and reality offered in the media report. Coupled with the wish of media reporting to garner a certain response it seems that media reports affect not only the way an individual approaches reality but it also aims for the receiving subject to return to consequent ways of thinking for further insight into reality, encouraging the consumer to buy more of its media project.

Both inter-subjective and subject-object relations are Pre-conditioned in the individuals taking part in the creation of a media report. In addition, their historicity, language, religion, upbringing, race, sex etc. all affect the way that the report is made. In a way each report is unique to the individuals in question. With all individuals subject to a commonality of experiences they are responsible for a multitude of varying responses. That is not inconceivable considering the uniqueness and variety of individuals involved, even if all the possible responses themselves might be innumerable. We therefore need to be aware of these things that affect us, as well as the way in which they in turn affect others. It is in these situations that it is important to understand the perspectives of those involved as well as what is being reported, how and why.

Bibliography:

Bødker, Henrik. "Muslims in Print, or Media Events as Nodes of Cultural Conflict." In *Media, Religion and Conflict*, edited by Lee Marsden and Heather Savigny. 81-94: Ashgate Publishing, 2009.

Couldry, Nick. "Media Rituals: From Durkheim on Religion to Jade Goody ". In *Exploring Religion and the Sacred in a Media Age*, edited by C. Deacy and E. Arweck. Ashgate, 2009.

Denham, B.E. "Building the Agenda and Adjusting the Frame: How the Dramatic Revelations of Lyle Alzado Impacted Mainstream Press Coverage of Anabolic Steroid Use." *Sociology of Sport Journal* 16, no. 1 (1999): 1-15.

Elmelund-Præstekjær, C., and C Wien. "Mediestormens Politiske Indflydelse Og Anatomi' [the Political Influence and Anatomy of the Media Storm]." *Nordicom Information* (2008): 19-28.

[15] Poulton, "English Media Representation of Football-related Disorder: 'Brutal, Short-hand and Simplifying'?," 32.

Ewing, Katherine Pratt "Living Islam in the Diaspora: Between Turkey and Germany." *South Atlantic Quaterly* 102, no. 2/3, Spring/Summer (2003): 405-31.

Hall, S. "The Treatment of Football Hooliganism in the Press." In *Football Hooliganism: The Wider Context*, edited by R. Ingham. London: Inter-Action Imprint, 1978.

Lang, G.E., and K Lang. *The Battle for Public Opinion: The President, the Press and the Polls During Watergate*. New York: Columbia University Press, 1983.

Loftus, Elizabeth F. "Leading Questions and the Eyewitness Report." *Cognitive Psychology* 7 (1975): 560-72.

Mantzavinos, Chris. *Naturalistic Hermeneutics* [in English]. Cambridge University Press, 2005.

Poulton, Emma. "English Media Representation of Football-Related Disorder: 'Brutal, Short-Hand and Simplifying'?". *Sport in Society: Cultures, Commerce, Media, Politics* 8, no. 1 (2005): 27-47.

Shadid, W. "Berichtgeving over Moslims En De Islam in De Westerse Media: Beeldvorming, Oorzaken En Alternatieve Strategieën." *Tijdschrift voor Communicatiewetenschap* 33, no. 4 (2005): 330-46.

Shadid, W., and P.S. van Koningsveld. "The Negative Image of Islam and Muslims in the West: Causes and Solutions." In *Religious Freedom and the Neutrality of the State: The Position of Islam in the European Union.*, edited by W. Shadid and P.S. van Koningsveld. 174-96. Leuven: Peeters, 2002.

Taylor, Charles. "Interpretation and the Sciences of Man." *Review of Metaphysics* 25, no. 1 (1971): 3-52.

P-16

Emerging from the cell: The role of the cytoskeleton in determining cell architecture

David Mentlak

Prof. P.J. Hussey, University of Durham

Summary: Cells are often displayed as perfectly spherical units. However this is not the case and cell shape, size and pattern varies greatly depending on the function of the cell. Projections outside of the cell are also common in nature and serve a variety of purposes. One of the best studied examples of cellular protrusions are those used in cell migration. Finger-like projections are crucial in guiding migrating cells in their environment and the actin cytoskeleton, a filamentous protein scaffold within the cell, is the major driving force behind these emergences. Understanding how the actin cytoskeleton and the proteins associated with it function in forming these projections is an important area of cell biology research and impacts upon our understanding of disease.

Introduction

The fundamental unit of all life is the cell. From "simple" unicellular bacteria to "complex" multicellular eukaryotic organisms such as animals, plants and fungi, they all share the same functional unit; the cell. Contrary to school textbooks, not all cells are either spherical in animals or cuboidal in plants. Rather, cells come in various different shapes and sizes depending on its functional role. For example, the biconcave disc shape of red blood cells provides a high surface area to volume ratio that maximizes gas exchange. Furthermore cellular architecture is not limited to smooth edges or straight lines; cellular protrusions are common across all forms of life and serve various important roles in cell biology. Cellular protrusions are highly dynamic structures that are able to assemble and disassemble through feedback systems involving the cytoskeleton, cell membrane and external signals. Elucidating the proteins that are involved in generating these structures is not only an important area of cell biology research but has further implications in understanding disease.

Emergences from the cell are common and can come in a wide variety of forms in the eukaryotic cell, with each fitting a particular function. Various stable projections exist in certain cell types for example cilia and microvilli. However the focus shall be on cellular protrusions that are more dynamic and can respond to the environment, specifically examining emergences involved with cell motility.

During cell migration (e.g. during immune responses, wound healing and embryonic development), two emergences form, driving cell motility. Firstly, a thin sheet emerges from the leading edge of the cell termed the lamellipodia. This first cellular protrusion is caused by the branched polymerisation of the actin cytoskeleton against the membrane forcing the leading edge forward. Secondly, finger-like projections, termed filopodia, also emerge from the lamellipodia at the leading edge of a migrating cell. Filopodia are also actin based structures but these are formed through the formation of tight actin bundles pushing at the cell membrane. Polymerisation of actin filaments against the cell membrane, in both lamellipodia and filopodia, drives this edge forward which provides the force needed for migrating cells (*1*). In order to understand how these emergences function and how they form it is necessary to briefly introduce the cytoskeleton and the contribution it makes to various cellular processes.

What is the cytoskeleton?

The cytoskeleton is an essential component of all cells and in animal cells is composed of three filamentous networks; actin microfilaments, microtubules and intermediate filaments. As filopodia are primarily created through actin filaments polymerising against the cell membrane, this element of the cytoskeleton shall be focussed on.

The actin cytoskeleton is involved with a large number of important cellular functions in animal cells. Cell motility, controlling cell shape, cell adhesion, organelle transport, cell-cell communication and

signal transduction are all governed by the actin cytoskeleton. The actin cytoskeleton is composed of globular subunits (G-actin) which can polymerise into long helical filamentous actin (F-actin) and depolymerise back, creating a dynamic filament. The G-actin subunits are asymmetric in structure and are all aligned in the same direction during polymerisation therefore creating a polar filament. The actin cytoskeleton can form a wide variety of structures within the cell and each arrangement relates to a different cellular function. As mentioned the actin cytoskeleton can form into tight bundles in filopodia, other examples include the branched network seen in lamellipodia and stress fibers which are used in cell migration.

The dynamicity of the actin cytoskeleton and the variety of structures it forms is not achieved solely through the individual protein filaments. Rather, a whole suite of accessory proteins that bind actin, termed actin binding proteins (ABPs), are able to influence its properties. Broadly speaking there are three classes of ABPs; those which influence actin dynamics (e.g. creating new filaments, severing filaments), those which use filaments as tracks for transport (e.g. myosin proteins, transporting cargo along actin filaments) and those which use actin as a scaffold (e.g. proteins that bind actin filaments and form higher order structures like tight bundles or loose gel like arrays, and also bind actin to other cellular compartments). Understanding how the different ABPs function and interact provides great insight into how the cytoskeleton is utilised in the cell. The spatial and temporal coordination of all the different classes of ABPs allows the actin cytoskeleton to form a wide variety of structures and participate in many different cellular processes.

What are cellular protrusions?

During cell migration, finger-like projections are produced from the lamellipodia (Fig. 1A) Filopodia perform a vital function during cell migration; they examine the microenvironment that the cell is moving into. The cell creates these emergences to test the environment for what is ahead as the filopodia contain receptors at the cell membrane which are able to transduce signals within the cell in response to external stimuli (*1*). Furthermore filopodia also contain adhesion molecules that are able to bind to the external environment (*2*). These initial adhesion sites, which are coupled to the actin cytoskeleton, can develop further into more secure sites which are essential for cell migration. Adhesion sites allow something for the contractile machinery of the cell (actin filaments in the form of stress fibers) to pull against.

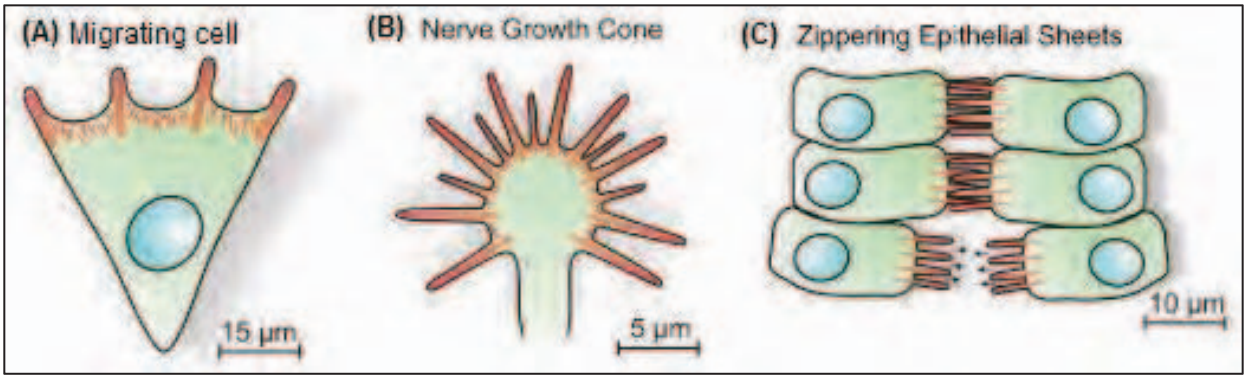


Fig 1. Filopodia in different cell types. (A) Thin filopodia emerge from the lamellipodia at the leading edge. (B) Neuronal growth cone. (C) 'Adhesion zippering' in an epithelial sheet. Diagram modified from (*3*).

Filopodia also have a function in wound healing, embryonic development and neuronal growth-cone pathfinding. The filopodia formed in epithelial cells (cells that form into sheets) during wound healing and embryonic development both share similarities in their behaviour. Firstly, filopodia form in these occasions as they are needed in cell motility when two epithelial sheets fuse. However, the cells in epithelial sheets can travel as a collective sheet of cells rather than solitary migrating cells. Here, the filopodia have the same function as in solitary cells as they probe the environment and help two matching sheets fuse together. Filopodia from opposing sheets interdigitate together and form

adhesion sites between the two sheets. The filopodia help to pull the two target sheets together in a process termed ‘adhesion zippering’ (Fig. 1C.) (*4*). The sites of filopodia contact then mature into strong cell-cell contacts.

The role of filopodia role as an environment probe is also essential for neuron development. Neuronal growth cones are motile structures of neurons that guide axons and dendrites to their target site (Fig. 1B). This is done through following particular chemicals and cues which filopodia have a role in sensing. Furthermore initial filopodia emergences have a role in forming more stable dendritic spines (*1*). These are an important structure at the synapse of some neurons and relay signals from one neuron to the other. The filopodia precursors of these structures can mature into the dendritic spines with the correct signal.

Cell motility clearly has a strong connection with the emergence of filopodia and this intrinsic behaviour can also be seen in malignant carcinoma cells. As carcinoma cells become metastic they have to migrate and therefore utilise filopodia. The number of filopodia on carcinoma cells also correlates with an increased invasiveness (*5*). Filopodia are utilised in cell motility along a planar substrate, however carcinoma cells need to invade through rigid tissues. Therefore carcinoma cells also contain invadopodia, another emergence with the ability to rearrange the external environment to migrate through 3D tissues (*6*). Understanding the molecular components that control migration and metastasis could provide novel targets for anti-cancer drugs.

Recent research has also identified a link between filopodia and autism spectrum disorders (ASD) (*7*). This work showed that a mutation in a scaffolding protein prevented filopodia from maturing into dendritic spines and was associated with ASD.

How does the cytoskeleton cause cellular protrusions?

Filopodia are created through the polymerisation of tight actin bundles against the plasma membrane thus causing a finger like projection from the cell. The polymerisation of actin against the plasma membrane, in both filopodia and lamellipodia, drives the migrating cell forwards (*1*). Many different molecular components that bind to the actin cytoskeleton are crucial in governing this polymerisation and mentioned below are some of these proteins involved.

ABPs involved in controlling actin dynamics play a key role in filopodia formation. ENA/VASP proteins are large proteins involved in filopodia formation in mammalian cells and localise to filopodia tips. These proteins appear to have an ‘anti-capping’ activity, meaning they prevent proteins from binding the end of filaments which prevent their elongation (*8*). Thus ENA/VASP allows the continual growth of the filopodia. Furthermore a variant of an ENA/VASP protein has recently been found to be overexpressed in certain breast cancers (*9*). ABPs controlling the creation of new actin filaments are also integral to filopodia formation. Formins are a class of actin nucleators which create an unbranched network of actin filaments (unlike the Arp2/3 complex, present in the lamellipodia, which creates actin filaments at a 70° angle to the existing filament). Unbranched nucleation is critical in creating long straight filaments and a specific type of formin, Dia2, has been shown to be recruited in filopodia formation (*1*).

Myosins, the ABPs that mediate transport activity along actin filaments also contribute to the formation of cellular protrusions. Recent studies have shown that increasing the expression of myosin X in various different cell types induces filopodia formation (*1*). This protein could possibly have a role in transporting other important components to the filopodia tips as it has the ability to interact with many other filopodia proteins.

The unbranched actin filaments formed in the filopodia need to be organised into tight parallel bundles, and therefore require ABPs that create higher order structures. Fascin is an actin bundling protein that can bind multiple filaments together thus strengthening the filopodia (*1*). Several epithelial cancers have increased levels of fascin which is correlated with increased metastasis and

invasiveness (*10*). Several other ABPs exhibit bundling activity and are present in the filopodia, including the ENA/VASP proteins (*1*).

In addition, other molecular components that don't bind actin are also needed for cellular protrusions to occur. I-BAR domain proteins have been shown to cause membrane deformation and thus facilitate the cytoskeleton in protruding through the cell (*1*).

Controversies exist over how all these molecular components come together to create filopodia, as the molecular machinery is not identical between every cell type and organism. However a recent model presented by Mattila & Lappalainen (2008) suggests the possible mechanism for emergences from the cell (Fig. 2). This model proposes that unbranched actin filaments or formin nucleated filaments converge at a point at the membrane, mediated by the activity of myosin X (Fig. 2A). These filaments then polymerise and force the membrane forwards at this point which could be facilitated by the membrane deforming I-BAR domain proteins (Fig. 2B). As these filaments elongate, they are bundled by fascin and possibly ENA/VASP proteins to become more stable structures (Fig. 2C). Myosin X could also have a role in transporting essential components along the filopodia to maintain its integrity.

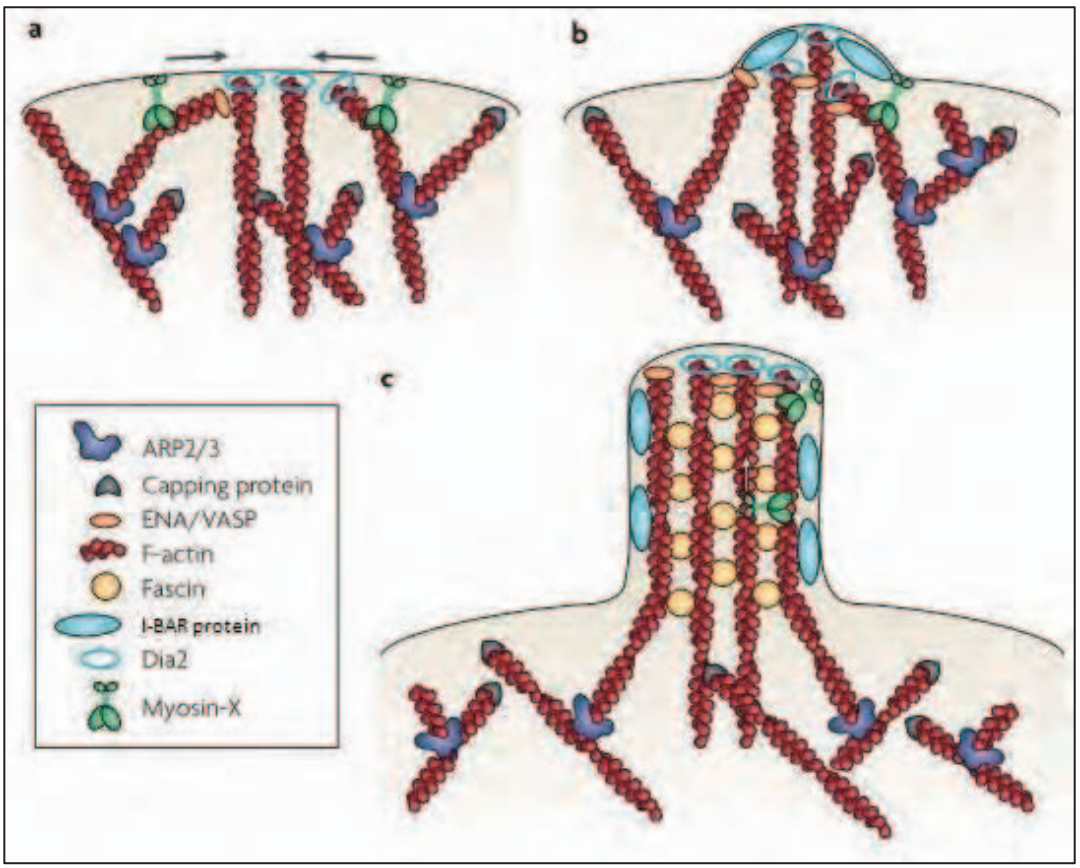


Figure 2. Model of filopodia formation. (A) Convergence of unbranched actin. (B) Polymerisation of actin causing the emergence. (C) Elongation of actin filaments and maintenance of filopodia. Diagram modified from Mattila & Lappalainen (2008).

The role of ABPs in governing filopodia formation and fully elucidating how they function as well as discovering new ABPs is critical for understanding this important area of cell biology. Furthermore, although not discussed here, the signalling mechanisms underlying the spatial and temporal regulation of these ABPs is crucially important.

Conclusion

Cellular emergences are common in various different cell types and can perform various different roles. Filopodia, the finger-like projections, seen in the leading edge of migrating cells have a key role in cell motility. The importance of these structures has an impact on various different cellular processes including cell migration, wound healing, development and the immune response. The actin cytoskeleton has a vital role in creating these structures as actin filaments polymerise against the cell membrane driving it forward. However, it is the orchestration of various other accessory proteins interacting with the actin cytoskeleton and the cell membrane that results in filopodia emerging from the cell. Controversies still exist over how filopodia form and project out of the cell but further research into the molecular mechanisms involved could elucidate how this process occurs. However inherent difficulties arise in unravelling filopodia formation as different cell types are not identical and therefore neither is the molecular machinery controlling it.

References: *1*. P. Mattila & P. Lappalainen, *Nature Reviews Mol. Cell Biol.* **9**, 446-454 (2008). *2*. B. Hoffman & C. Schäfer, *Cell Adh Migr*. **4**, 190–193. (2010). *3*. J. Faix., D. Breitsprecher, T.E. Stradal, and K. Rottner *Int J Biochem Cell Biol* **41**:1656–1664 (2009). *4*. S. Khurana & S. George, *Cell Adh Migr.* **5,** 409-420 (2011). *5.* D. Vignjevic., et al., *Cancer Res*. **67**, 6844–6853 (2007). *6*. H. Yamaguchi, & J. Condeelis, *Biochimica et biophysica acta*. **5**, 642-652 (2007). *7.* S. Berkel, et al. *Hum. Mol. Gen.* **2**, 344-357 (2012). *8*. J. Bear & F. Gertler, *J. Cell Sci.,* **122**, 1947-1953. *9*. S. Goswami, et al., *Clin Exp Metastasis*, **26**, 153-159 (2009). *10*. L.M., Machesky, & A., Li. *Commun. Integr. Biol*. **3**, 263-270 (2010).

P-17

The emerging patterns and problems of climate change: what can we learn from arctic plants?

Rachael Oakenfull
Durham University, Biological and Biomedical Sciences

Summary: The arctic is one of the harshest environments for plant growth in the world. As climate change progresses the climate in each environment will alter significantly, causing more frequent occurrences of extreme weather such as, freezing temperatures, drought and floods. Food worldwide is already limited and as these extreme weather events become more frequent, more and more crops will be destroyed. By studying three species of arctic *Vaccinium* (close relatives of commercial blueberries and cranberries), we aim to further understand the mechanisms behind plant freezing tolerance to identify potential targets for crop improvement to increase sustainability, potentially minimising the damage climate change can cause.

1. Introduction: By 2050 the global demand for food will have doubled it is estimated that an extra 1-2 billion people will not have enough to eat due to climate change. Already in 2012 someone starves to death every 3.6.seconds.

Climate change is defined as a long term alteration in the usual climate of an area which can include unexpectedly low temperatures, floods and drought. Each time there are occurrences of altered or extreme weather conditions valuable food crops are destroyed. Temperature is already responsible for destroying vast areas of food crops globally each year and is a major factor in explaining the size and distribution of plants. In recent years there have been a number of examples worldwide of crops being lost to freezing temperatures, in April 2012 multiple states in America lost fruit crops, In 2011 Mexico has lost 90% of its maize crops to frost, which in turn caused food prices to triple. In the same year India had its coldest winter in 30 years and Zimbabwe (a major exporter of flowers) lost a large number of flower crops to frost.

Plants usually survive winter by acclimating, triggered by temperatures slowly becoming lower and a decreasing day length. This combination of light and temperature signals to the plant to acclimate as winter is approaching, allowing the plant time to prepare internally, by altering their cellular and metabolic processes to allow survival of freezing temperatures. Depending on the species the acclimation process can take between a few days to several weeks to be complete.

Freezing damages plants when the water inside the cell walls of the plant turns to ice, this causes water to move out of the cell causing dehydration inside the cell. Similarly, this also occurs when water between cells freezes. The major cause of plant damage from freezing is caused by dehydration either through water loss due to the cell membranes freezing or through losing water throughout the day to the air surrounding them. When the ground freezes, the roots are unable to take up water causing the plant to dehydrate. Severe dehydration will kill the plant.

As climate change progresses sudden freezing events will become a more common occurrence and not all plants will be able to survive the change in temperature. Freezing events caused by climate change cause problems as their occurrence is becoming more random so the plants will not have the warning period of slowly decreasing temperatures in which to acclimate. Another set of problems relate to winter warming events where a period of warm weather over winter can cause plants to de-acclimate and even bud burst as in spring time then when the freezing temperatures return, the plant is no longer prepared and the flower or stem buds are badly damaged by freezing. finally, some plants such as tomatoes are sensitive to freezing temperatures and no matter how long they are given to acclimate they will not be able to prepare to the level where they can survive freezing temperatures. Already we have seen changing weather patterns in the UK with unusually cold spells earlier in the year and sudden warm spells in winter.

The arctic covers around 4million km^2 with temperatures, snow cover and light levels varying depending on the exact position. For this research the region of the arctic referred to is Abisko. Situated in the north of Sweden, 200km north of the Arctic Circle. Temperatures fluctuate between -35^{o}C and +18^{o}C over the year allowing areas of permafrost to melt for a short 2-3month period over summer. There are large open boggy areas consisting of dwarf shrubs, mosses and lichens called tundra. Some of the same plants can be found in Abisko and the UK but the arctic species are dwarf due to their harsh environment. In arctic conditions it is advantageous for plants to remain small to stay below the snow level where temperatures are around -10^{o}C in winter, whereas outside of the snows protection it can be -35^{o}C.

By studying 3 arctic species of *Vaccinium* (*Vaccinium myrtillus, Vaccinium uliginosum, Vaccinium vitis-idaea*) we can learn more about the mechanisms behind how plants are adapted to and survive in extreme cold conditions. The berries from two of these species are the edible fruit lingonberry and bilberry and are closely related to cranberries and blueberries therefore adding a commercial interest. These three species were used as they are all found in the same environment but have different freezing tolerances making them ideal to compare the molecular mechanisms of freezing tolerance. *Vaccinium myrtillus* (*V. myrtillus*) is deciduous with green stems, soft leaves and a freezing tolerance of around -10^{o}C, living in sheltered areas under snow or beneath tree cover. *Vaccinium uliginosum* (*V. uliginosum*) is deciduous with wooden stems, tougher leaves and a freezing tolerance of between -10^{o}C and -40^{o}C growing in both sheltered areas and exposed tundra. *Vaccinium vitis-idaea* (*V. vitis-idaea*) has tough green stems, tough waxy leaves and a freezing tolerance of -40^{o}C growing in exposed rocky outcrops without snow cover in winter.

In the model organism *Arabidopsis thaliana* it has been found that a CBF transcription factor is essential for freezing tolerance. CBF behaves like a switch in response to low temperatures which switches on a large number of genes responsible for acclimation and therefore freezing

tolerance. In *Arabidopsis* it can switch on the acclimation process within 15 minutes

Results: The CBF sequences cloned from the 3 *Vaccinium* species were compared to see if the differences in freezing tolerance were due to differences in protein sequence (molecular differences) or caused by their different physical characteristics. (Cloning is essentially just copying a piece of DNA from an organism). The protein and DNA sequences in each species differed to each other and contained huge differences to the *Arabidopsis* sequences. Each CBF was then cloned and put into *Arabidopsis* plants to compare the effects caused by just the new CBF. These new hybrid plants were then tested for the effects of the new CBF on other processes in the plant, including freezing tolerance and their general appearance as growth response (dwarfing) is part of the cold acclimation process.

The plants containing the CBF from *V. myrtillus* were smaller than usual with very flat leaves that are much darker than the usual appearance of *Arabidopsis* and also different to the other hybrids which looked normal. In experiments like these plants appearing smaller is usually a sign of increased freezing tolerance as just like in the arctic a large amount of energy is used preparing the plant to be freezing tolerant leaving less energy available for growth resulting in smaller plants. The hybrid *Arabidopsis* was then tested for freezing tolerance. Showing that the *Arabidopsis* with the *V. myrtillus* gene had an increased level of freezing tolerance without acclimation in comparison to the other hybrids and also the normal *Arabidopsis* plants.

2. Discussion: The increased freezing tolerance in the *V. myrtillus* hybrid plants futher highlight that CBF is an important part of acclimation in some plants, making it an important target to study in crop plants. However, the negative results from the *V. uliginosum* and *V. vitis-idaea* suggest that there may be other target genes which in freezing tolerant plants have a greater effect. The next logical step is therefore to look for further target genes in *Vaccinium,* before taking this on to looking at their effect on crop plants. The fact that the freezing tolerance in Arabidopsis can be increased by adding a single gene gene from *V. myrtillus* gives very positive hopes for the future.

References:

Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T., and Thomashow, M.F. (2001). Components of the Arabidopsis C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in Brassica napus and other plant species. Plant Physiol 127: 910-917.

Polashock, J.J., Arora, R., Peng, Y.H., Naik, D., and Rowland, L.J. (2010). Functional Identification of a C-repeat Binding Factor Transcriptional Activator from Blueberry Associated with Cold Acclimation and Freezing Tolerance. J Am Soc Hortic Sci 135: 40-48.

Sonesson M, Callaghan TV (1991) Strategies of Survival in Plants of the Fennoscandian Tundra. Arctic 44: 95-105.

Wang, Q.J., Xu, K.Y., Tong, Z.G., Wang, S.H., Gao, Z.H., Zhang, J.Y., Zong, C.W., Qiao, Y.S., and Zhang, Z. (2010). Characterization of a new dehydration responsive element binding factor in central arctic cowberry. Plant Cell Tiss Org 101: 211-219.

Nov.28 (oral presentation)

08

Physicists behaving badly
Or: (Why) does physics play a natural role in interdisciplinary studies of emergence?

Tom McLeish
Durham University

Summary: Physics is the most transgressive science; it keeps trespassing on other disciplines' territories. This bad behaviour can annoy people, but also responds to a very general methodological power in the discipline, which is a science uniquely non-defined by any field or object of study. In this brief paper I examine progressively bold excursions away from physics' traditional home-grounds, keeping the idea of emergent phenomena as a motif, and beginning in the currently "emergent" area of Soft Matter Physics.

1. Introduction : Sciences typically define themselves by their field of enquiry. *Geo*logy, *Astro*nomy, *Bio*logy, *Paleo*ntology, *Chemi*stry, and so on, make the point. This is not true of physics, whose etymology is much more general: φυσις is usually translated "nature". The name knows no boundaries, and increasingly the science has behaved accordingly. From the mid-19th century fragmentation of our current scientific disciplines away from the older, all-encompassing "natural philosophy", physics has ramified from its original core fields of matter, energy, electromagnetism and atomic structure. Etymologies track this re-emergence of a general science for us as well: for some reason communities of physicists have demarked sub-disciplines within new territory that their methods have begun to investigate: astrophysics, geophysics, biophysics, chemical physics. A new wave of questions in biology is beginning to attract the name "biological physics" to distinguish it from the mid-20th century expedition that spawned "biophysics". Are these aggressive strategies of an invading discipline, establishing the academic equivalent of pallisaded beachheads on foreign soil? Or do they represent negotiated boundary treaties, enshrining the promise that physics will go thus far but no further into the heartland of other disciplines? The question is urgent, for the physical and life sciences are not the only lands that physics has more recently sought to colonise. The arrival of "econophysics" and even "sociophysics" signals a late 20th century move towards the social sciences. Should we be concerned about this ambitious sister-discipline, or welcome her ambition?

We might in any case seek to understand what happens when physics (in the form of course of a community of people who call themselves "physicists") approaches fields that normally carry other labels. The topic of this symposium, "emergence", is a very promising starting point. For it exposes the methodologies that physics deploys in making new topics its own. Physics is a master of analogy: "force", "coupling", "interaction", "collective behavior", "correlation", "field" are metaphors that apply to materials as well as to magnetism, to evolving populations of organisms, even to crowd behavior. The key to understanding the promiscuity of physics is perhaps to be found in the unique way that it uses mathematics to breathe life into its metaphors. All science since the enlightenment has arguably been dominantly reductionist, and physics is not an exception. But by making measurements on many scales, and by harnessing the power of mathematics to connect the imagined interactions of the fine-grained constituent parts of a system to its large-scale behaviour, it provides an embodiment of emergence.

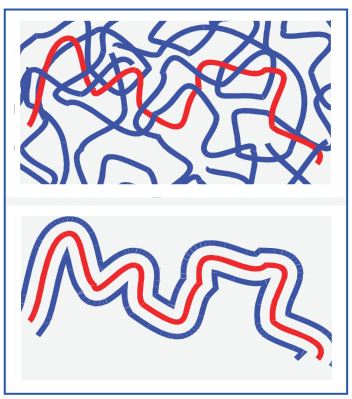
In the following we will examine in a little more detail some examples of "transgressive physics" that each seek to uncover an emergent phenomenon.

2. Discussion: We begin with a series of examples that can be charted in terms of both the physics and the sociology of academia. In all cases, a very successful application of physical methods, thinking, modeling, abstraction or coarse-graining, met with strong opposition from at least some parts of the disciplinary area that laid prior claim to the topic of investigation.

Physics meets Chemical Engineering: Entangled Polymer Dynamics

One of the greatest, possibly the greatest contribution of the story of science to human learning to date is the molecular basis of matter. Nobel Laureate Richard Feynmann famously identified it as the one "single-sentence" encapsulation of discovery that he would advocate passing on to the seeds of a new civilization having to rise from the self-destructive ashes of our own, "that all matter is composed of small particles, constantly in motion, that attract each other when far apart and repel when very close". This single idea, conjectured of course by the ancients[i], has huge, explanatory power, allowing us to grasp the reasons for melting, evaporation, boiling and many other properties of matter. Lucretius and his commentators had already noticed the possible macroscopic consequences of "atoms" endowed with different shapes; Simplicius gives us a famous passage in which the different fluid properties of oil and wine are explained by the relative "slipperiness" of their constituent atoms. These ancient ideas received a distant echo in the 1970s when a remarkable theory was proposed by Masao Doi and Sam Edwards in Cambridge[ii]. Doi and Edwards sought to account at the molecular level for the extraordinary properties of molten polymeric ("plastic") fluids. These are liquids when flowing slowly, but under rapid deformations behave more like elastic solids. It was already known that the molecules constituting these materials are very unusual: they have the form of very long and extremely flexible strings of atoms linked together, continually twisting and coiling under thermal agitation.

The consequences of such random dynamics of stringlike entities in isolation were relatively well understood, thanks to theoretical and experimental work on dilute solutions of polymer molecules in the 1950s and 60s. The challenge was to understand the behaviour as concentration is increased and the chains begin to overlap, until in the melt they fill all space. The simplest experiment that can be made on the macroscopic material is a small deformation accompanied by a measurement of the force (stress) supported by the fluid. These "viscoelastic" fluids show a measurably finite time in which the induced elastic stress falls to zero. As a solution is made more concentrated (once the molecular chains start to overlap), or if samples of higher chain length are made, this "relaxation time" increases dramatically. The viscosity of the fluid increases even more rapidly, together with a suite of non-linear effects that appear only at high deformations. But why?



The imaginative guess was that, at high overlap, it is the *topological* constraints on the chains that dominate. Simply, two strings cannot pass through each other. At high density any one chain is highly constrained from motion perpendicular to its own contour, and only free to move along it, much like a snake caught in a tube, a motion termed "reptation" from the snake analogy. The idea is pictured on the left. The system of many chains caricatured in the upper panel is conceptually replaced by an *effective* constraint on any given chain – the tubelike region in the lower panel. The *ansatz* is simple to put into mathematical terms, the coarse-grained dynamics

of any chain is essentially one-dimensional (along its contour), and from this follows the scaling of viscosity and relaxation time with chain length and concentration. It is even possible to deduce some of the non-linear properties of the melt with no recourse to any free parameters.

We note immediately that the entire physical system is emergent. The effective tubes, and the resultant motion of the polymers, quite different from its nature in dilute solution, are self-generated because of the interactions. It would in principle be possible for the chains to adopt configurations that did not "entangle" with each other in such a dramatic way, but this is prevented by the overwhelming statistical likelihood that the chains adopt random walk configurations for most of the time. A fundamental derivation of the emergent tube-structures has proved elusive for more than 30 years. One reason for this is that there is no "perturbative" approach to this problem. The tubes are no more the property of two interacting chains than they are of one. It is only the very large number of overlapping chains (it is simple to show that this quantity scales as $N^{1/2}$, where N measures the chain length). Only recently has it been possible to carry out very large numerical simulations that capture the formation of the effective tubes as a function of very many mutual topological interactions of chains.

The original idea was disarmingly simple, and disarmingly successful, yet highly controversial. One problem was that models for viscoelastic fluids had been until then the declared territory of chemical engineering. I recall in my early academic life the most bitter confrontations at conferences. Fuelled by the successes of the model on one side, and by the data it failed at first to account for on the other, it was clearly also born of a deep-seated disciplinary clash. The physicists claimed the first truly molecular theory of this type of matter, the engineers (perhaps surprisingly) charged the newcomers with a lack of rigour. Because computer simulations at that point were not large enough to capture the emergent structures, they were sometimes taken as evidence that the new ideas were wrong, carrying titles such as *The Death of Reptation*. But the beautiful idea, grounded in experience with "self-consistent fields", rather natural to theoretical physics in the 20th century, proved to be the fruitful idea that the engineering questions needed. 30 years later it is the basis for industrial design protocols for new plastic materials.

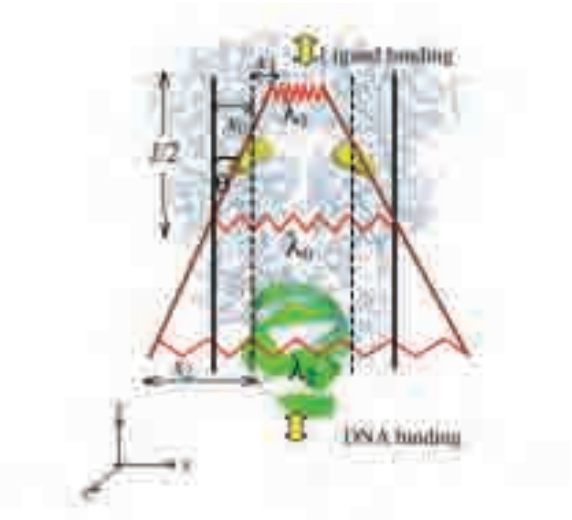
Physics meets Molecular Biology: Protein Thermodynamics

The life sciences command a field of very different history, methodology and epistemology to that of the physical sciences, as several detailed meta-studies have recently shown[iii]. This in spite of the common parlance of "scientific method" with the implication of a universal set of guiding principles and practice. There is no space here to go into the details of distinctions, but others have identified and evidenced the very different roles in physics and biology played by: hypotheses, counter-factuals, coarse-graining, models, the constraints of physical law, and mathematics. Each one creates potential pitfalls for collaborative ventures that bring the methodologies and drivers of physics into the biological world. The idea of emergent behaviour, though initially one that would seem natural within the mindset of biological science (for after all, are not all super-molecular structures in biology, from cells to tissues to organisms "emergent" from their constituent parts in a strong sense?), is one where the contested nature of methodologies can become explicit.

In this exemplar we focus on the twin challenges of coarse-graining and physical law in the emergence of thermodynamically-generated correlations in protein-binding to small molecules. There are two areas of "trespass" here for the physicist. The first is the notion that at larger length-scales, the details of the many specific and local interactions between atoms can be averaged over to generate

emergent and correlated dynamical structures that are not dependent on the very local details. Protein biochemistry has (for many good reasons) focused on the role of specific, highly evolved molecular detail in molecular structure that, for example, determines the degree of "fit" of the binding pockets of large protein molecules to the smaller signaling molecules that bind to them. The second challenge is the compromise of structure. Since the extraordinary success of early X-ray protein crystallography in determining the mean structure of DNA and example proteins, the idea that "structure determines function" has been almost a mantra of molecular biology. But physics knows that molecular material at length-scales of a few nanometers, and at temperatures of 300K is subject to large thermal fluctuations that randomly distort the relative geometry of atomic positions very significantly, and on timescales of a nanosecond. There is no escape from this constraint of thermodynamics: evolved biology cannot escape it. Therefore it must be able to recruit it.

These insights were behind a programme of work to model and measure the avenues provided by large-scale (on the scale of a protein molecule in entirety) fluctuations on the cooperativity of binding. Ubiquitously in molecular biology, the presence of one small molecule bound to a site on a large protein macromolecule affects the tendency to bind of another small molecule at a distant site. Such "allosteric" communication is canonically explained by a shift in structure. The problem is that in many cases there is *no* observed change in mean structure at all, so the "structure-to-function" paradigm fails. However, the thermal fluctuations that themselves compromise or modulate the structure turn out to provide alternative channels of allosteric communication. If there exist modes of fluctuation by which an entire protein "breathes" coherently, then a restriction on any one site modulates the entire mode, so affecting in turn any other potential binding site on the protein. Such a statement can be cast in the mathematical language of statistical mechanics, yielding quantitative estimates of the free energy changes available for allosteric signalling. Such a calculation was first performed at a coarse-grained level for a model of the *lac* repressor protein dimer. Each of the two sizeable "monomers"[1] was modeled as a rigid block, with the interface between them a looser elastic domain represented by nine locally-tunable harmonic potential wells [iv] (illustrated in the figure). Although successfully published in the physics literature (in the top physics letters journal), this work drew the ire of some molecular biologists, revealed in some anonymity-preserving fieldwork of biology-physics collaborations[v]: "I was singularly unimpressed – we know that real proteins are nothing like this model" (attributed to "Professor C").



When pressed, the objection stemmed from a combination of contrasting views of epistemology and methodology. The methods of detailed X-ray crystallography and NMR structural

[1] in biochemical language "monomer" refers to a potentially large protein in its own right, complexed with one or more others, contrasting with the normal chemical usage of a small molecular block concatenated with identical copies in a polymer.

analysis reveal an atomistically-resolved knowledge of protein structure. To a structural biologist, a protein "is" the data represented on the international "protein data base" and its linked repositories. Each unit (or "residue") of the protein has its own local substructure and interaction with its neighbours. Furthermore, it is characterized by its own, again local, degree of motion, revealed by NMR spectroscopy. A less-well spatially resolved notion of a protein looks like a retrograde step, a loss of information and a concomitant loss of knowledge. The total of protein physics is simply the sum of the parts in this view. The problem is that such knowledge, while appearing detailed and complete, actually misses essential information about *emergent* properties of the whole structure. Even at the level of a harmonic "spring-like" energy function for the momenta, p and coordinates, x of the residues,

$$H = \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{p} + \mathbf{x}^{\mathrm{T}}\mathbb{K}\mathbf{x}$$

the free energy of the protein contains information on correlated dynamics that in turn create the emergent "information channels" through the complex, exploited by signaling mechanisms to regulate binding with no apparent change to the structure itself. Taking the partition function from H, the evaluation appears at the level of its determinant – so independent of any particular representation and emphasizing the essential connection between the dynamical normal modes of the protein (for these diagonalise H, and turn Z into a simple product:

$$Z = \int \cdots \int dx_1 dx_2 \ldots dx_n \exp\left(-\frac{H(x_1, x_2, \cdots, x_n)}{k_B T}\right) = \frac{(2\pi k_B T)^n}{|\mathbb{K}|^{\frac{1}{2}}}$$

Calculating the free energy $F=-k_B T\log Z$ from this partition function plainly releases an allosteric free energy of order $k_B T$ for every global normal mode of motion. A major challenge for experiment is a technique that picks out the long-range correlations in motion underpinning the signaling channel. A remarkable aspect to this short tale is that the emergent property, and its character as emergent, is really one of relationship between the human scientist and the object (Strang, this volume) rather than an ontological entity of itself (there is no non-linearity, coupling, dissipation in the underlying physics). Emergence sometimes requires nothing more than recognition to do work for us.

Physics meets Sociology: Sociophysics and Popularity (the Zipf distribution)

In one of the more extraordinary ventures of physics in general, and of statistical mechanics in particular, is the attempt to address social behaviour in terms of stochastic yet predictive models. The newly-termed "sociophysics" has its roots in an earlier, and less controversial (for unclear reasons) movement in quantitative economics. Irving Fisher, professor of economics at Yale, was the student of J.W. Gibbs, a pioneer of statistical mechanics in the 19th century. The idea of a "statistical mechanics" of populations was famously fictionalized by Issac Asimov in his "Foundation" trilogy, playing on an analogy with the ergodic assumption of statistical physics: a sufficiently large population of stochastic and interacting entities explores the space of their possible configurations in a representative way, so that fluctuations become small relative to a predictable mean, or expected, behaviour.

Recently promoted by Serge Galen and others[vi], the subtitle, "do humans behave like atoms?" invites the obvious and immediate answer "no, of course not!" But when Galen used a computer model of group opinion-dynamics to predict the outcome of both French and Dutch

referenda on the European constitution, the question took on more nuance. Might, in some circumstances, the emergent behaviour of internally-communicating groups of individuals, be predictable? The predictions would carry the same caveats as those of statistical physics of molecular matter: they cannot be exact, because of the fundamentally stochastic nature of the underlying dynamics. However, when populations sizes become large, or the mean behaviour involves an average over very large numbers of agents, or when the fundamental dynamical rules can be simplified, the results can become surprisingly robust. The spread of opinion through small-group interaction that Galen modeled in the cases of the referenda, drew on two of these criteria. The voting population of individuals was of course tiny compared to the huge numbers of degrees of freedom encountered in material physics, but still large under a fluctuation criterion that estimates error as the square root of the population size. The other assumption, redolent of the successful methodology of theoretical physics to keep the minimum assumptions absolutely necessary, was a simple rule of small-group dynamics with a slight bias to the *status quo ante*[2].

Some numerical observations of emergent phenomena in human systems almost demand an underlying set of universal rules. An example is found in the statistics of popularity, encapsulated in "Zipf's law"[vii]. The frequency f of occurrence of an item (word frequency, song-playing popularity, population size of cities, …) is plotted as a function of rank order, n, (with $n=1$ the most popular, and so on). Many of these distributions follow to very close approximation the power law $f(n)\sim n^{-k}$ with $k=1$. There have been a number of attempted explanations of the law, but in the spirit of sociophysics it seems appropriate to investigate a minimal model, and proceed to test it, before setting up more detailed hypotheses. It will prove another informative example of emergence of unexpected global patterns from simple local rules.

We observe that, underneath a popularity distribution is a dynamical system. Rank order is not static for any individual item (*e.g.* word usage is changing: neologisms are invented, grow and spread, while old usages die away). Items are continually exploring trajectories of popularity, driven by the stochastic elements particular to any system (in the case of words, invention, conversation, imitation, broadcasting). The stochastic nature of the sum of all these dynamical processes must induce a diffusive structure onto the probability frequency function $P(f)$ which describes the number of items occurring with frequency f (note that this is a different representation of the same information as Zipf's law described in terms of rank order, and that $f(n)\sim n^{-k}$ implies that $P(f)\sim f^{1-1/k}$). individual items diffuse up or down in popularity space as attempted "sharings" of that item succeed or fail. Care must just be taken with the value of the diffusion constant locally to the frequency f in the distribution. The greater the frequency of occurrence of an item in a population of individuals, the larger gains it stands to make if successfully adopted, and the larger falls it risks by an average reduction in usage. We will begin with a general case of the effective diffusion constant $D(f)=D_0 f^{\mu}$. So

---

[2] The simulation rule is easy to state. Agents hold one to two opinions, one for change and one for the *status quo ante*. A single dynamical step involves clustering them into small groups numbering from three to ten. Each agent emerges with the majority opinion of the group, except in even numbered groups where opinions tie. In this case all emerge with the opinion of the *status quo ante*.

$$\partial f/\partial t = D_0 \quad \partial / \partial n \, (n^{\mu} \, \partial f/\partial n)$$

Looking for stationary solutions we find those satisfied straightforwardly by $P(f) \sim f^{1-\mu}$. Now let us consider appropriate values for the nonlinear diffusion exponent, μ. In a single timestep for interactions of a popularity group of f items there will be a number proportional to f of communications. In each case the expectation value of Δf, the mean step in frequency is also of order f. Since for small diffusive intervals $\Delta f^2 = D\Delta t$ this induces an additional quadratic dependence of D on f leading to the result $\mu = 3$. In turn this leads directly to the Zipf distribution.

Implications of socio-physics for the understanding of both individual and social behaviour have yet to be mapped. An initial corollary is that less weight might be attributed to "intelligent" or "computed" reactions than has traditionally been assumed (this is already known to be a flaw in models of economics that assume agents compute, and then act on, the choices that optimize their financial benefits[viii]). However, as in the biological example, there is even greater risk here that at best important details, and at worse the entire causal complex, are being overlooked, in an over-reached exercise in coarse-graining.

Conclusions

Physics draws on a combination of mathematical, conceptual and epistemological roots that enable it to conceptualise and recreate emergent phenomena. Furthermore, the process naturally challenges traditional disciplinary boundaries. When it does so it needs to tread, and to listen, carefully.

References

[i] E.g. Lucretius, *De Rerum Naturae*

[ii] M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics*, Oxford, (1986)

[iii] T.C.B. McLeish "Physics met Biology, and the Consequence was…", in *Studies in History and Philosophy of Biological and Biomedical Sciences*, Ed. D. P. Rowbottom, *Studies in History and Philosophy of Science C*, **42**(2), 190-192 (2011).

[iv] R.J. Hawkins and T.C.B. McLeish, "Coarse-Grained Model of Entropic Allostery", *Phys. Rev. Letts*, **93**, 098104 (2004)

[v] D. P. Rowbottom, in *Studies in History and Philosophy of Biological and Biomedical Sciences*, Ed. D. P. Rowbottom, *Studies in History and Philosophy of Science C*, **42**(2), 190-192 (2011).

[vi] S. Galen, *Sociophysics*, Springer (2012)

[vii] G.K. Zipf, *Human behaviour and the principle of least effort*, Addison-Wesley (1949)

[viii] *E.g.* P. Mirowski, *Machine Dreams*, CUP, 2002

09

Softmatter Physics : Multiscale simulations for polymer melt flow

Takashi Taniguchi
Chemical Engineering, Graduate School of Engineering, Kyoto University

Summary:
We have developed a multiscale simulation technique that combines microscopic polymer dynamics simulations and a macroscopic flow simulation to investigate the history-dependent flow behavior of polymer melt. In the multiscale simulation, a smoothed particle hydrodynamics method is used to solve the macroscopic flow and a Langevin stochastic model is used to solve the dynamics of entangled polymer chains in each fluid element. We apply the multiscale simulation technique to polymer melt flow passing a circular object in a two-dimensional periodic system. It is found that the strain-rate history-dependent stress of the polymer melt affects its flow behavior, and this stress memory causes nonlinear behavior even in the region where the Weissenberg number is less than unity. The multiscale simulation method enables us to evaluate the macroscopic spatial distribution of microscopic quantities, such as the spatial distribution of the number of entanglement.

1. Introduction

Products made of polymeric materials have been indispensable for our daily life. One of the important characteristics of polymeric material is the thermo-plasticity that it can be easily molded at a high temperature and processed by controlling its state, which is beneficial in a variety of practical applications. It has been known that the melt of polymeric material and their mixtures can exhibit a wide range of flow behaviors depending on the dynamic response of the polymer's microscopic internal states under an imposed strain or strain-rate [1]. Especially, melts of polymeric materials used in industrial applications usually exhibit non-Newtonian flow behaviors (cf. Newtonian fluids: a local stress of the fluid is simply proportional to the local strain-rate at the same time and space). So, the prediction of flow behaviors of polymeric liquids using computer simulations have been a challenging theme in various fields of science and engineering including physical science, materials science, and mechanical and chemical engineering. Polymeric liquids are known to exhibit peculiar flow behaviors, e.g., a rod climbing flow (known as the Weissenberg effect), die swelling (the Barus effect), that are related to the micro-scale dynamics and state of their polymer chains; these dynamics affect the viscoelasticity, shear thinning/thickening behaviors, and flow induced phase transitions of these liquids [1].

For non-Newtonian fluids, many novel computational fluid dynamics (CFD) schemes have been proposed for dealing with the viscoelastic flows of polymeric liquids [2-6]. In CFD methods, a model constitutive equation [7-14] is necessary to determine the local stresses at each instant from the history of previous velocity fields. However, the detailed forms of the constitutive relations for polymeric liquids (melts) composed of polymer chains with arbitrary polymer architecture are generally unknown [15]. Even for polymeric liquid consists of chains having simple branching architectures, the expression is mathematically quite complicate. Thus, the usual CFD methods, which require

constitutive equations, are usually not straightforwardly applicable to the complicated flow problems of polymeric liquids. On the other hand, microscopic simulations such as MD and CG simulations are often used to investigate the rheological properties of such materials. There is a merit to use microscopic simulations because one can easily incorporate any microscopic molecular pictures into the model. The microscopic simulations are usually performed for a tiny piece of the material in the equilibrium or non-equilibrium state under uniform external fields of shear velocity, temperature gradient, and electric field. Although the microscopic simulations are even applicable to macroscopic flow behaviors, the drawback of this type of simulation is the enormous computation time required [16]. Thus, for problems that concern large-scale and long-time fluid motions far beyond the molecular scale, which are commonly encountered in engineering, fully microscopic simulations are difficult from a practical standpoint. One of the ways to overcome the weaknesses of the individual methods is to develop a multiscale simulation method that combines CFD and MD [or CG] simulations.

When we develop such a multiscale simulation method, however, we easily face difficulties that are not only the enormous computational time but also the physical validity and consistency of the method. If the characteristic time scales in microscopic and macroscopic dynamics are quite different, one can assume that a microscopic state always reaches a local steady state at every instant time under an external condition which is macroscopically determined, and then the computational time would be reduced by using sophisticated techniques which are developed for investigations of equilibrium state. However, as the length of polymer chain is larger, the characteristic times of the micro-scale dynamics of polymer chains tend to be very long and are often comparable to the time scales of macro-scale fluid motions. Thus, for these compounds, one cannot separate the micro- and macro-scale dynamics in the temporal domain; one cannot make the assumption usually made for simple fluids under flow that the fluid elements are in local steady states at each instance. Therefore, one must trace the "memory" or "history" of the deformations of each fluid element along its streamline. This coupling between micro- and macro-scale dynamics hinders the simulation of polymeric liquids. Therefore, it is difficult in general to predict a flow of polymeric fluid due to this tight coupling between a local macroscopic flow and the microscopic internal state in the same place. Actually, when a polymer melt consists of polymers with a molecular weight M higher than the entanglement molecular weight M_e the polymer molecules are entangled, and the relaxation time of the polymer conformation is long compared to that of dilute polymer solutions because of the entanglement. Therefore, even for a homogeneous bulk polymer melt, it is very difficult to predict the rheological behavior by molecular dynamics simulations without any coarse-graining procedures accounting for the microscopic internal degrees of freedom. On the other hand, the reptation theory [7-9] has widely been used to predict rheological properties in bulk of polymeric liquid composed of linear polymer chains. In the model,

the dynamics of well-entangled polymer chains are treated as a dynamics of a chain in a tube made of other polymer chains in a mean field manner. The original reptation theory [7-9] and the extended ones [10-14] have succeeded in explaining the complex behavior observed in many experiment of polymer melts. However, it is difficult to apply the theory to arbitrary polymer melts because it is too mathematically difficult to incorporate molecular pictures of polymer with arbitrary architectures into the extended reptation theories [10-15] which are basically described by a Fokker-Planck equation. On the other hand, Langevin stochastic models [17-20] have been developed and reproduced the rheological properties of various polymeric materials with branches and/or molecular weight distributions [17, 18]. These models use efficient numerical computation to predict the bulk rheology of polymer melts; however, they are not applicable to macroscopically inhomogeneous flows. The macroscopic flow behavior of polymer melt is usually predicted using a fluid dynamics simulation with a constitutive equation that describes a nonlinear relation between the stress and strain-rate fields without microscopic details, as mentioned before. No general constitutive equation that is applicable to arbitrary polymer melts exists because a general polymer melt has extremely complex molecular states, which makes the stress response unpredictable even in a simple flow pattern. As described above, each of the microscopic and macroscopic approaches has limited applications.

As an alternative to using a constitutive equation, here we propose a simulation technique that incorporates Langevin course-grained simulators into a fluid dynamics simulation. This approach is a type of micro-macro simulation [21-25] that was developed to consider history-dependent flow behaviors of polymer solutions. Our approach uses the Lagrangian fluid particle method to solve the macroscopic fluid dynamics, and therefore the dependence of the stress on the history of the strain-rate is directly involved [26-28] without any special treatments which are necessary in connecting Eulerian and Lagrangian schemes [21]. When the polymer solution consists of long molecules without any entanglements, the bulk behavior of polymer solution can be treated as a single-body problem, while the entangled polymer melt is a fully many-body problem. Therefore, a fully Eulerian technique such as the Brownian configuration field method [24, 25] is difficult to apply to entangled polymer melt system since there exist difficulties in treating macroscopic convection of entangled polymers. In flows of well-entangled polymer melts, the convection of microscopic internal states is intrinsically important to describe their flow behaviors since the conformations of polymers at microscopic level cause not only linear response but also non-linear and/or retarded response such as shear-thinning and stress over shoot. However, in considering a flow with symmetry such as translational or rotational symmetry [29-31], the fully Eulerian ones are still useful even for polymer melts.

In my presentation, I give a detailed explanation of a multiscale method that bridges the hydrodynamic motions of fluids using computational fluid dynamics (CFD) and the microscopic [or mesoscopic] dynamics of polymer configurations using molecular-dynamics (MD) [or coarse-grained

(CG) Langevin-type] simulations. The concept of bridging micro- and macro-scale dynamics is also important for other flow problems of softmatters with complex internal degrees of freedom (e.g., colloidal dispersion, liquid crystal and so on). Although the basic idea of our multiscale method is applicable to those softmatter flows, by focusing on a well-entangled polymer melt, we demonstrate how our proposed multiscale simulation works and to show physical information that can be obtained only in this multiscale simulation method.

2. Multi-scale simulation using Lagrangian fluid particle method

In our multiscale simulation method, to maintain the microscopic information such as entanglements and deformation in polymers, we use a Lagrangian fluid particle simulation technique in which each fluid particle has a microscopic level simulator that accounts for the internal states of the fluid particle. Assuming that the polymer melt is isothermal and incompressible fluid, the governing equations for fluid particles which constitute polymer melts are given by the following equations:

$$\rho_i \frac{d\boldsymbol{v}_i}{dt} = -\nabla p_i + \nabla\cdot\boldsymbol{\sigma}_i + \boldsymbol{F}_{bi} \qquad \frac{d\boldsymbol{r}_i}{dt} = \boldsymbol{v}_i \tag{1}$$

$$\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i(\boldsymbol{\kappa}_i) \tag{2}$$

$$p_i \equiv p_i(\{\rho_i\}) \tag{3}$$

where the velocity gradient tensor $\boldsymbol{\kappa}_i$ at the position of i-th particle is defined as $\boldsymbol{\kappa}_i \equiv {}^{\mathrm{T}}(\nabla\boldsymbol{v})_i$, and $\boldsymbol{F}_{\mathrm{b}}$ is a body force. These equations are solved by using macroscopic variables except for Eq. (2). The local stress tensor $\boldsymbol{\sigma}_i$ can be evaluated from microscopic variables describing polymer dynamics under an instantaneous local strain-rate $\boldsymbol{\kappa}_i$. The slip-link model [17, 18] is one of simulation models which can nicely describe the dynamics of polymer chains. The model is composed of confining tubes with some entanglement points called as slip-links which confine a pair of chains and represent effective constraints in virtual space. Averaged number of slip-links or entanglements on a chain is represented as $Z \equiv M/M_{\mathrm{e}}$. In the simulation, we trace the configurations of confining tubes constrained by the slip-links. The slip-links are relatively convected each other and the confining tubes are deformed according to the macroscopically obtained local velocity gradient tensor $\boldsymbol{\kappa}$. The reptations of polymer chains generate or eliminate slip-links. For given configurations of the chains, the stress tensor $\boldsymbol{\sigma}^{\mathrm{p}}$ coming from a deformation of entangled polymer chains is calculated by

$$\sigma^{\mathrm{p}}_{\alpha\beta} = \sigma_{\mathrm{e}} \sum_i \left\langle r^{\mathrm{s}}_{i\alpha} r^{\mathrm{s}}_{i\beta} / |\boldsymbol{r}^{\mathrm{s}}_i| \right\rangle / a_{\mathrm{s}}$$

where a_{s} is a unit length of slip-link model and $r^{\mathrm{s}}_{i\alpha}$ is an α-component of the i-th tube segment vector (α=x, y, or z) connecting between adjacent slip-links along a chain. The unit of stress σ_{e} in the slip-link model relates to the plateau modulus G_{N} as $\sigma_{\mathrm{e}} = (15/4)G_{\mathrm{N}}$ [17]. The slip-link model has two characteristic time-scales; Rouse relaxation time τ_{R} and the longest relaxation time τ_{d}. Rouse

relaxation time τ_R and the longest relaxation time τ_d relate to Z as $\tau_R = Z^2 t_e$ and $\tau_d \propto Z^{3.4} t_e$ [9, 17], where t_e is the time unit of the slip-link model. The relaxation of contour length of confining tube occurs on the time-scale of τ_R, while that of orientation occurs on the time-scale τ_d. These two characteristic times appear in the stress relaxation. All polymer simulators on the fluid particles maintain polymer configurations at every time steps; and the saved configurations are used as the initial conditions at the next time step. Usually, macroscopic time unit t_{macro} and microscopic time unit t_{micro} have a large time-scale gap, and therefore it is needed to divide the macroscopic time unit t_{macro} into Nt_{micro}. Since the slip-link model used here is sufficiently coarse-grained and the time unit t_e can be treated as the same time-scale of macroscopic fluid t_{macro}, we employ $t_{macro} = t_{micro} \equiv t_e$. Note that we set $\boldsymbol{\sigma} = \boldsymbol{\sigma}^p + \boldsymbol{\sigma}^d$ for local stress of macroscopic fluid where $\boldsymbol{\sigma}^d$ is an extra dissipative stress tensor. Since the slip-link model is a Langevin type coarse grained model based on the reptation theory, microscopic dynamics of less than tube segment size is treated as a random force exerted to a slip-link, and the contribution from the microscopic dynamics does not explicitly appear in the stress tensor of the slip-link model. We assume the dissipative stress $\boldsymbol{\sigma}^d$ to be the Newtonian viscosity $\eta_d \mathbf{D}$, where $\mathbf{D} \equiv \boldsymbol{\kappa} + {}^T\boldsymbol{\kappa}$ is a rate of strain tensor.

As mentioned in Introduction, the local stress tensor that is needed to solve the macroscopic flow can be evaluated from microscopic variables describing polymer dynamics under an instantaneous local strain-rate. As shown in Fig. 1, the fluid system is composed of a large number of fluid particles with the radius *a.* The flow behavior is described by the motions of the set of fluid particles.

The main procedures of the simulation are summarized briefly as follows:

(i) Calculate the local density and the strain rate tensor at the position of each Lagrangian particle in SPM manner at the macroscopic level.

(ii) Obtain the pressure from the density distribution at the macroscopic level.

(iii) Update the local slip-link model of the *i*-th fluid particle under the local strain-rate at the position and then obtain the local stress tensor from the resulting configuration of the slip-link model. This procedure is executed on each fluid element in turn.

(iv) Calculate the force coming from stress tensor and the pressure at the macroscopic level.

(v) Update the velocity and position of all the fluid particles by using the equation (1).

(vi) Return to the item (i).

The local microscopic dynamics of polymer chain (i.e., the slip-link simulations) in (iii) is composed of the following physical processes: (P1) affine deformations of the medium, (P2) contour length fluctuations and reptation motion of polymer chains and (P3) the constraint renewal of polymer chains. Details of (P1)–(P3) are as follows. (P1) The tube segment $\boldsymbol{r}_j^s$ is affinely deformed according to the equation $d\boldsymbol{r}_j^s / dt = \boldsymbol{\kappa} \cdot \boldsymbol{r}_j^s$. (P2) Contour length fluctuations and reptation motion are implemented in the evolution for the contour length L of the polymer chain:

$$dL/dt = -(L(t) - L_{\mathrm{eq}})/\tau_{\mathrm{R}} + \dot{L}_{\mathrm{affine}} + \sqrt{2a_{\mathrm{s}}^2/3t_{\mathrm{e}}Z\Delta t}\; g\,,$$

where $L_{\mathrm{eq}} \equiv Za_{\mathrm{s}}$ is the mean equilibrium length, $\dot{L}_{\mathrm{affine}}$ is the change in L due to (P1), and g is a Gaussian random variable with zero mean and unit variance. (P3) The constraint renewal process updates the number of entanglements Z as follows. The number of entanglements Z is changed by the motion of the chain ends. When the length of a chain end tube segment becomes zero, the entanglement node just on the chain end is removed. Simultaneously, the node coupled to the removed entanglement, associated with a different chain, is also removed mimicking the constraint release. On the other hand, when the length of the chain end tube segment becomes larger than a_{s}, a new entanglement node is created on it, and a random vector is selected for the new end segment. At the same time, a new node is created on another randomly selected polymer chain and coupled to the just created node at the chain end, which mimics constraint creation. The further detailed explanation on the slip-link model is given in ref. [17].

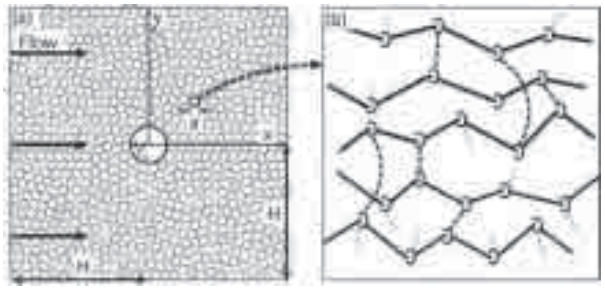


Figure 1 – (a) Polymeric flow around a cylindrical obstacle with a radius r_{c}. The small circles represent Lagrangian fluid elements with radius a, each of which have a microscopic simulator schematically shown in (b). (b) Schematic view of our used Langevin type model describing well-entangled polymer melts. A ring denotes a slip-link describing entanglement point and the dotted line in (b) means the two chains are entangled at the point.

Integrating Eq. (1), we update $\{\mathbf{v}_i\}$, $\{\mathbf{r}_i\}$. From a new configuration of particles, we calculate the density at the position of each particle in similar fashion to the usual smoothed particle hydrodynamics [32]: $\rho_i = \Sigma_j m_j W(|\mathbf{r}_j - \mathbf{r}_i|;\, h)$, where m_i is mass of i-th particle and $W(|\mathbf{r}|;\, h)$ is a Gaussian shape function with cutoff length $2h$. The deviation of local density from an initially given constant density ρ_0 results in the local pressure force $-\nabla p$. To obtain the spatial derivative of the velocity field, stress field, and pressure field ($\nabla\boldsymbol{\upsilon}$, $\nabla\cdot\boldsymbol{\sigma}$, ∇p), we use a technique which is developed in the modified smoothed particle hydrodynamics [33, 34].

It should be noted that our multiscale model can be shown to be consistent with "GENERIC" [35, 36], a modern framework for non-equilibrium thermodynamics, when we consider the time evolution of the internal energy [36]. The GENERIC formalism can be rather effective while using our multiscale model to study mesoscopic problems where thermal fluctuations are important. However,

since we assume the present system to be isothermal, we do not consider the flow of the internal energy but focus our attention on how the microscopic internal states affect the macroscopic fluid behavior. This treatment is indeed less rigorous but quite sufficient for our present purpose.

By using the above algorithm, we consider a polymer melt flowing in the *x*-direction and passing an infinitely long cylinder oriented in the *z*-direction with a radius r_c. Because of the symmetry of the system, we can treat the system as two-dimensional, and the flow can be described as two-dimensional flow in the *xy*-plane. We assume a non-slip boundary condition for the velocity on the surface of the cylinder and periodic boundary conditions at the boundaries of the system. As a microscopic simulator we used the slip-link model that can accurately describe the dynamics of entangled polymers.

The dimensionless parameters governing the present problem are the Reynolds number $\mathbf{Re}=\rho U r_c/\eta^0$, the Weissenberg number $\mathbf{Wi}=\tau_d D_{xy}$, and the shear viscosity ratio η_d/η^0, where U is the average flow velocity. The zero shear viscosity η^0 of the polymer melt is given by $\eta_p{}^0+\eta_d$, where $\eta_p{}^0$ is the zero shear viscosity of a polymer melt described only by the slip-link model. From a rheological data derived from the slip-link simulation with average Z=7 entanglements in the bulk, we obtain the longest relaxation time $\tau_d \simeq 200 t_e$ and the zero shear viscosity $\eta_p{}^0 \simeq 17.5 \sigma_e t_e$ where t_e is the time unit of the slip-link model. The cylinder radius r_c was set to $3a$, where a is the unit length in the fluid particle simulation, and we assign the unit mass m to all fluid particles. The wall of the cylinder consists of fixed fluid particles evenly spaced on the perimeter. Each fluid particle consists of 10,000 polymers; enough to describe the bulk rheological properties of the polymer melt under an imposed shear and/or extensional deformation. About 1000 fluid particles with 10,000 polymers each are evenly placed in the system; therefore, we need to simultaneously solve the dynamics of ca. 10,000,000 polymers. Each slip-link simulation can be performed independently of the others, making parallel computing effective in this multiscale simulation. In order to induce a flow around the obstacle, the fluid is subject to the external field described by the body force $\boldsymbol{F}_b=(5\times10^{-4},0)$ $(\eta_0/t_e a)$, the flow becomes steady-state in about $1000t_e$. The average flow velocity in steady-state in this system is nearly equal to $(0.04,\ 0)a/t_e$ for a fully Newtonian flow $\eta^0 = \eta^d$ (entanglement stress=0), and $(0.055,\ 0)a/t_e$ for a polymer melt flow with Z = 7 and η_d/η^0= 0.1. In both cases, **Re** is less than 0.2, and the flow is laminar. In the polymer melt case, De is higher than 3.6, and the flow depends on the flow history. The average flow velocity of the polymer melt is higher than that of the Newtonian flow, i.e., the shear-thinning behavior can be observed in the flow because the local Weissenberg number $\mathbf{Wi}\equiv\mathbf{Wi}(\boldsymbol{r})=\tau_d|D_{xy}(\boldsymbol{r})|$ is expected to be higher than unity in the vicinity of the cylinder.

To investigate the velocity field $\boldsymbol{v}$, strain-rate D_{xy}, and shear stress σ_{xy} in the steady state, we employ a linear interpolation of the data at the particle positions to determine the values at regular lattice points and then time-average the data evaluated at the lattice points. To reduce the noise in the data, the time averaging was carried out between $10^3 t_e$ and $2\times10^3 t_e$.

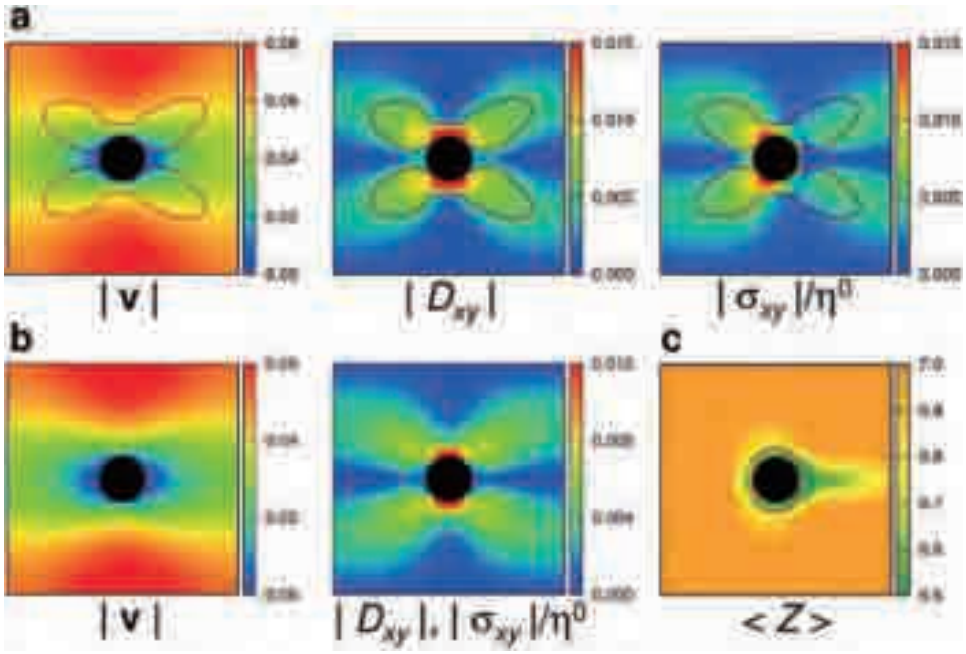


Figure 2 –: Two-dimensional flow around an obstacle. Color contour maps represent time-averaged magnitudes of the velocity field $|\boldsymbol{v}|$ $[a/t_e]$, strain rate $|D_{xy}|$ $[1/t_e]$, and shear stress over the zero shear viscosity $|\sigma_{xy}|/\eta_0$ $[1/t_e]$ for (a) the polymer melt with $\langle Z\rangle_{eq} = 7$ and $\eta_d/\eta_0 = 0.1$ and (b) the Newtonian fluid with $\eta_0 = \eta_d$ in a steady state. Panel (c) shows $\langle Z\rangle$ in the polymer melt. The regions inside the red lines in (a) correspond to **Wi**>1 and those in (c) to $\tau_R\dot{\gamma} > 1$, where $\dot{\gamma} \equiv \sqrt{\mathrm{Tr}(\kappa^T\kappa)}$.

Figure 2 shows the spatial distributions of $|\boldsymbol{v}|$, $|D_{xy}|$ and $|\sigma_{xy}|/\eta_0$ in the steady state for the polymer melt with $\langle Z\rangle_{eq}=7$ and $\eta_d/\eta_0 = 0.1$ and the Newtonian fluid with $\eta_0 = \eta_d$ (Fig. 2(b)). Unlike σ_{xy}, the magnitudes of $\boldsymbol{v}$ and D_{xy} in Fig. 2 appear to be nearly symmetric between the upstream and downstream regions and reflect the laminar nature of the flow. In Fig. 2(a), a nonlinear relationship between σ_{xy} and D_{xy} can be observed near the cylinder because in this region **Wi**>1 and consequently the melt shear-thins. Moreover, the shear stress σ_{xy} of the polymer melt exhibits an apparent asymmetry between the upstream and downstream regions. If the asymmetry in the stress distribution is caused only by the nonlinear relationship, the stress distribution should be symmetric in the regions where **Wi**≤1. However, the asymmetry is also observed in the regions where **Wi**≤1. Judging from the analyses of memory effect (that is not shown here) we conclude that the asymmetry in the stress distribution is caused by the history-dependent stress in the polymer melt.

Finally, in order to extract a relationship between microscopic state and macroscopic flow from the present multiscale simulation, we investigate $\langle Z\rangle$ the spatial distribution of entanglements. As shown in Fig.2 (c), we find that in the vicinity of the cylinder, the number of entanglements per chain slightly decreases . When $\tau_R(d\gamma/dt) \gg 1$ (τ_R : Rouse relaxation time, and $d\gamma/dt$: strength of a local shear rate), the contour length of the polymer is significantly longer than its equilibrium value, $a\langle Z\rangle_{eq}$. The chain

retraction process which attempts to return the chain contour length to its equilibrium value causes disentanglement. In the upstream and downstream regions around the cylinder, the reduction of $\langle Z \rangle$ is also observed because of nonzero $|D_{yy}|$. The Advection due to the flow broadens the region where $\langle Z \rangle < \langle Z \rangle_{eq}$ in the downstream from the cylinder. The tail length in the downstream region can be estimated to be $\tau_R U \sim r_c$. The spatial distribution of $\langle Z \rangle$ is quite similar to that of Tr(σ) [37, 38], because the number of entanglements depends on the elongation of polymers in this multiscale simulation. We hope that the innovative advances in the micro-rheology techniques will enable the observation of the spatial distribution of entanglements in the future.

3. Conclusions

We have developed a new multiscale simulation method to investigate history-dependent flow behavior in polymer melt. This multiscale simulation consists of a smoothed particle hydrodynamics simulation and a slip-link model describing microscopic dynamics of polymer chains in each fluid element. The multiscale simulation technique has been applied to polymer melt flow passing a circular object in a two dimensional periodic system to demonstrate the effects of history-dependent stress. We have found that the strain rate history-dependent stress in the polymer melt affects its flow behavior, and the stress field is retarded from the strain-rate field. The presented multiscale simulation is applicable to various polymer melts, because the slip-link model or the alternative course-grained models can address a variety of polymer architectures, e.g., linear and/or branched polymers, polymer blends, and poly-dispersed polymers. This multiscale simulation is advantageous because it employs a fully Lagrangian method at the macroscopic level, while conventional Eulerian-based techniques which have difficulties accounting for the macroscopic advection of microscopic internal states.

Reference :

[1] R. B. Bird, R. C. Armstrong, and O. Hassager, "*Dynamics of polymeric liquids*" Vol. **1** (John Wiley & Sons, New York, 1987).

[2] X. L. Luo and R. I. Tanner, *J. Non-Newtonian Fluid Mech*. **21**, 179 (1986).

[3] X. L. Luo and R. I. Tanner, *J. Non-Newtonian Fluid Mech*. **22**, 61 (1986).

[4] D. Rajagopalan, R.C. Armstrong, and R.A. Brown, *J. Non-Newtonian Fluid Mech*. 36, 159 (1990).

[5] R. Guenette and M. Fortin, *J. Non-Newtonian Fluid Mech*. 60, 27 (1995).

[6] P. T. Baaijens, *J. Non-Newtonian Fluid Mech*. **79**, 361 (1998).

[7] S. F. Edwards, *Proc. Phys. Soc*. **92**, 9 (1967).

[8] P. G. de Gennes, *J. Chem. Phys*. 55, 572 (1971);

[9] M. Doi and S. F. Edwards, *The theory of polymer dynamics*, (Oxford University Press, New York, 1986).

[10] D. W. Mead, R. G. Larson and M. Doi, *Macromolecule* **31**, 7895 (1998).

[11] G. Ianniruberto and G. Marrucci, *J. Rheol.* **45**, 1305 (2001).

[12] G. Ianniruberto and G. Marrucci, *J. Non-Newton. Fluid Mech.* **102**, 383 (2002).

[13] A. E. Likhtman and R. S. Graham, *J. Non-Newton. Fluid Mech.* **114**, 1 (2003).

[14] T.C.B. McLeish, and R.G. Larson. *J. Rheol.*, **42**, 81, (1998).

[15] R. G. Larson, *Constitutive equations for polymer melts and solutions,* (Butterworths, Boston, 1988).

[16] A. E. Likhtman, S. K. Sukumaran and J. Ramirez, *Macromolecule* **40**, 6748 (2007).

[17] M. Doi; J. Takimoto, *Philos. Tras. R. Soc. London, Ser. A*, **361**, 641 (2003).

[18] Y. Masubuchi et al., *J. Chem. Phys.*, **115**, 4387 (2001).

[19] A. E. Likhtman, *Macromolecule*, **38**, 6128, (2005).

[20] T. Uneyama, *J. Soc. Rheol. Jpn.*, **39**, 135, (2011).

[21] M. Laso; H. Öttinger, *J. Non-Newtonian Fluid Mech.*, **47**, 1 (1993);

[22] K. Feigl, M. Laso, and H.C. Öttinger, *Macromolecules* **28**, 3261 (1995);

[23] M. Laso, M. Picasso, H. C. Öttinger, *AIChE J.* **43**, 877 (1997).

[24] M. A. Hulsen, A. P. G. van Heel and B. H. A. A. van den Brule, *J. Non-Newtonian Fluid Mech.*, **70,** 79 (1997).

[25] P. Halin, G. Lielens, R. Keunings, V. Legat, *J. Non-Newtonian Fluid Mech.*, **79,** 387 (1998).

[26] T. Murashima; T. Taniguchi, *Europhysics Letters*, **96**, 18002 (2011).

[27] T. Murashima; T. Taniguchi, *J. Polym. Sci. B*, **48**, 886 (2010).

[28] T. Murashima; T. Taniguchi, *J. Phys. Soc. Jpn.* **81**, SA013 (2012).

[29] S. Yasuda and R. Yamamoto, Phys. Fluids 20, 113101 (2008).

[30] S. Yasuda and R. Yamamoto, Europhys. Lett. 86, 18002 (2009).

[31] S. Yasuda and R. Yamamoto, Phys. Rev. E 81, 036308 (2010).

[32] Monaghan J. J., *Rep. Prog. Phys.*, **68,** 1703 (2005).

[33] G. M. Zhang and R. C. Batra, *Comput. Mech.*, **34**, 137, (2004).

[34] M. B. Liu, W. P. Xie, and G. R. Liu, *Appl. Math. Model*, **29**, 1252, (2005).

[35] H. C. Öttinger, Beyond Equilibrium Thermodynamics (Wiley-Interscience) 2005.

[36] M. Ellero, P. Español and E. G. Flekkøy, *Phys. Rev. E,* **68**, 041504 (2003).

[37] Rasmussen H. K. and Hassager O., *J. Non-Newtonian Fluid Mech.*, **56**, 65 (1995).

[38] Chilcott M. D. and Rallison J. M., *J. Non-Newtonian Fluid Mech.*, **29**, 381(1988).

10

Tipping points and emergence in the behaviour of landslides

David Petley
Department of Geography, Durham University

Summary: Mass movements are frequently used as a model to illustrate emergent behavior associated with self-organising criticality, and there has frequently been an assumption that real systems behave like simple sandpiles. However, in reality landslide systems are much more complex than the simple sandpile models, with many interacting factors, such as the variable presence of pore water, interparticle bonding and changes in external triggering mechanisms. This paper will explore tipping points and emergence in three contexts – landslide initiation, landslide size (volume and area) relationships and landslide impacts (losses of life).

1. Introduction: Magnitude – frequency relationships in mass movements are frequently cited as an example of emergent behavior. Indeed, the original proposal of self-organising criticality by Bak *et al.* (1987) used a sandpile in which grains were added by slow sprinkling to trigger new debris avalanches as its underpinning analogy. This in turn drove a wave of modeling of the failure behaviour of granular systems with the most famous being the Oslo rice pile model. This has led to an assumption by some that landslides display similar emergent behavior. However, this is not necessarily the case as terrestrial landslide systems differ considerably from those modeled in sandpile experiments. Critical differences include the dynamic setting (slopes that have evolved rather than being formed through sprinkling); the presence of pore fluids (especially water), which dominate the triggering of failure by altering the stress state; complex internal structures; and the presence of interparticle bonding. Thus, it is not immediately clear that real slopes will show the types of behavior displayed in sandpile experiments. This paper will explore the degree to which this is the case.

2. Landslide initiation: More than 95% of landslides occur as a result of an external trigger (forcing) such as heavy rainfall or seismic excitation. The key triggering mechanism is through a change in the stress state within the slope, causing an increased shear stress and/or a decreased normal stress state. Thus, behavior differs importantly from the sandpile model, where slopes fail because of the addition of new material, which then causes a cascading effect. Natural slopes essentially behave in one of two ways, dominated by either brittle or ductile behavior in the landslide mass. In the case of brittle materials, failure occurs through the formation of a shear surface – essentially the propagation of a crack through the landslide mass. In this case, the development of final failure is driven by the late stages of crack growth through the landslide mass. Thus, although the final rapid movement event appears to be spontaneous, measurement of strain shows that it is a smooth (albeit rapid) transition from pre-failure creep. This has allows successful and in some cases reliable prediction of the time of final failure for such slopes. The point of initiation of rapid movement is in a sense a tipping point in terms of overall behavior (a transition from partially unsheared to totally sheared materials). However, for slopes dominated by ductile behavior, failure is dominated by flow type processes best modeled by model that incorporates state- and rate-dependent friction. As such rapid movement events are much more complex to predict, and the presence of tipping point behavior is less obvious.

3. Magnitude – frequency relationships in landslides: The advent of high resolution satellite imagery and automated mapping technologies in recent years has allowed the collection of high quality inventories of multiple landslide events in the aftermath of large earthquakes and intense precipitation events, especially landfalling tropical cyclones, over large areas. Interestingly, the magnitude-frequency data for these landslides consistently show a fat-tailed power law distribution over a number of orders of magnitude in terms of both landslide surface area and volume. Perhaps surprisingly, there is little variation in the nature of the distribution between geographic areas, rock types and triggers. It is not entirely clear why this is the case at present. Interestingly, the relationship also breaks down at small landslide areas / volumes, and indeed at very small sizes it "rolls over", such that the numbers of events starts to decrease as the size decreases. There has been considerable speculation in the literature as to why this style of behaviour is observed. Initially it was suggested that this is a sampling problem – i.e. that it is difficult to collect representative inventories at the limits of the resolution of the remote sensing technique. More recently it has been suggested that this might be a rather more fundamental shift in behavior associated with a transition in dominant material type from rock (in larger landslides) to soil (which forms a thin mantle on the earth's surface, and thus is only relevant in shallow landslides).

4. Magnitude – frequency relationships in landslide losses: Interestingly, data on the loss of life from landslides also shows a fat tailed power law distribution, based upon data collected by the author since 2002, when analysed in terms of magnitude-frequency. In this case however the nature of the relationship varies between continents. Emergence in this case probably arises from two aspects – first, the emergent style behavior displayed by the landslides that cause loss of life, as described above. Second, social data, such as number of people per household, are also frequently described by fat tailed power law distributions. Thus, the combination of the physical and social systems is likely to promote this type of relationship, although note that the key underlying social factor is not clear. This could be the number of people per household, or it could be the number of households per square kilemetre, or both for example. In this context the differences between geographical areas (for example between India and China) are interesting. Rural areas of China in particular are undergoing rapid change at present, most notably perhaps with a key alteration in household size as outmigration to work and live in urban areas continues. Thus, it will be interesting to see whether the nature of the relationship between magnitude (i.e. numbers of lives lost) and frequency changes with time over the coming years.

References:

Bak, P., Tang, C. and Wiesenfeld, K. (1987). "Self-organized criticality: an explanation of 1/f noise". **Physical Review Letters** *59* (4): 381–384.

11

Landslides induced by typhoon Talas 2011 and long-term landscape development

Masahiro Chigira
Disaster Prevention Research Institute,
Kyoto University

Summary: Typhoon Talas 2011 hit the Japanese Islands and induced more than 70 deep-seated catastrophic landslides, causing severe damage in the Kii Peninsula. Landslides induced by the rainfall caused direct damage to houses, slid into swollen rivers and caused tsunami waves that moved upstream or rose up the opposite slope of the valley, and caused landslide dams, of which five persisted for more than half a year without being breached. The landslides resulted in 56 fatalities, with a further 32 deaths due to flooding or unspecified hazards in the Kii Mountains. These landslides occurred on slopes that had been formed by young river incision of paleosurfaces (old topographic surfaces). In addition, it was found that the landslides had been preceded by preparatory gravitational slope deformation, which appeared as small scarps along the head of landslides. The typhoon-induced landslides are thus ongoing phenomena to evolve landscape.

1. Introduction: Catastrophic landslides, which have been induced by rainstorms and earthquakes, are sudden, rapid, and long runout landslides, devastating wide areas commonly within a few minutes. They are very short time phenomena, but from another view, they occur as a process to evolve landscape in a geological time scale. I would like to present an example of catastrophic landslides induced by 2011 typhoon Talas in Japan; they occurred on slopes that had been gravitationally deformed in advance, and those slopes were located on slopes that had been newly made by young incision of rivers in paleosurface (old topographic surface). These topographic changes are geological phenomena, but also are important to predict potential sites of catastrophic landslides, which bring huge disaster to human society. Topographic features of catastrophic landslides before landsliding have scarcely been studied in detail before because usually aerial photographs have been used to observe topographic features, but 10 landslides induced by typhoon Talas had been surveyed by airborne laser scanner and had high resolution topographic data with 1-m mesh.

2. Landslides induced by typhoon Talas:

2.1 General features of rain and landslides

Typhoon Talas, which hit the Kii Mountains between 2 and 5 August 2011, caused more than 70 large and catastrophic landslides, estimated to be larger than 100,000 m^3 in volume (Ishizuka et al., 2012; Figs. 1 and 2). The rainfall associated with the typhoon fell between 31 August and 4 September, and exceeded 1000 mm over two thirds of the Kii Peninsula, while 2439 mm was recorded at

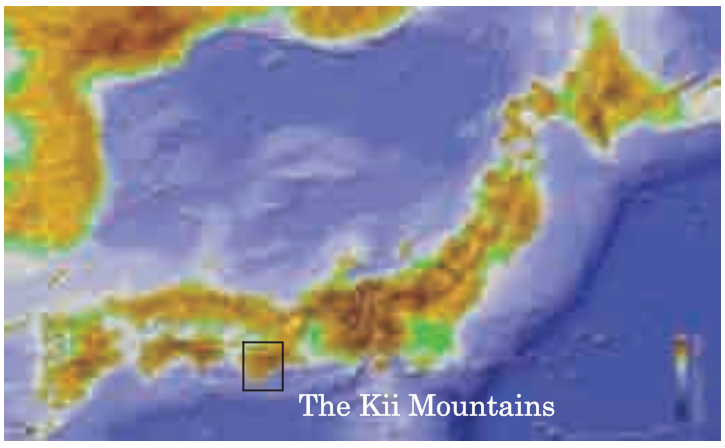


Fig. 1 Index map of the Kii Mountains.

Kami-Kitayama (Fig. 3). The rainfall amounts were quite large in comparison with previous precipitation in the Kii Mountains, where the average annual precipitation is 1300 mm in the northwest of the region, and over 3000 mm in the southeast (based on the 30-year record of the Automated Meteorological Data Acquisition System of the Japan Meteorological Agency). Despite such heavy rainfall over a wide area caused by typhoon Talas, large landslides only occurred in a rather sporadic manner, and shallow landslides did not occur widely, which is a similar pattern to the landslides associated with typhoon NABI (0514, Japan Meteorological Agency) in 2005, in the same Shimanto Belt in Kyushu (Chigira, 2009).



Fig. 2. Large landslides caused by typhoon Talas in the Kii Mountains. A: Akatani. B: Nagatono. C: Shimizu. D: Tsubonouchi (right: A; left: C). E: Kitamata. F: Iya.

The rainfall-induced landslides were concentrated in the center of the Kii Peninsula, which was not the area of highest rainfall (Fig. 3). This could be attributed to the variations in the pattern of rainfall and geological conditions. The Kii Mountains, located in the center of the Kii Peninsula, are 100 km wide and 100 km long, have a maximum elevation of 1915 m in the northeast, and are incised by the Kumano-gawa and the Kitayama-gawa; both flowing north to south (kawa or gawa means “river” in Japanese; Fig. 3).

Landslides induced by the rainfall caused direct damage to houses, slid into swollen rivers and caused tsunami waves that moved upstream or rose up the opposite slope of the valley, and caused landslide dams (Fig. 2).

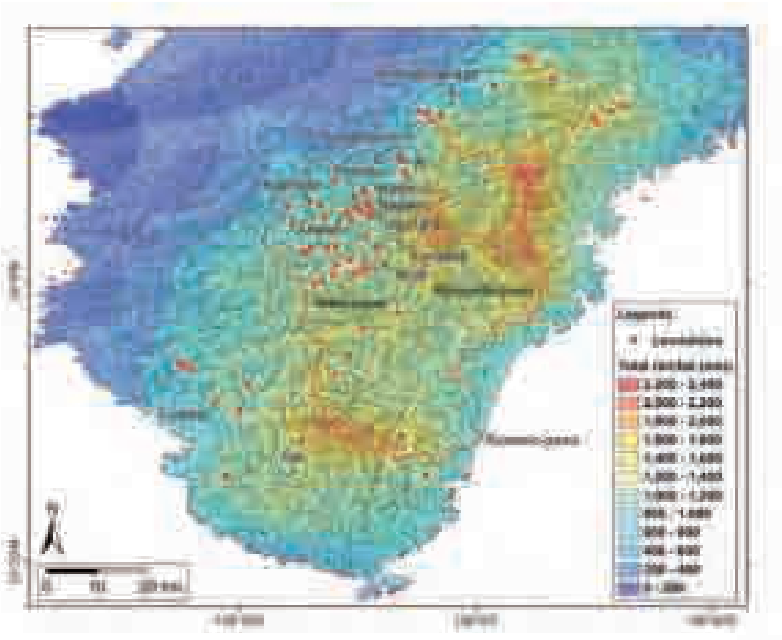


Fig. 3 Distribution of landslides and cumulative rainfall between 31 August and 5 September 2011.

The landslides resulted in 56 fatalities, with a further 32 deaths due to flooding or unspecified hazards in the Kii Mountains. Landslides started to occur at 6 p.m. (Japan Standard Time) on 3 September 2011 at Nojiri (Fig. 3), and continued until the evening of the next day. Landslide dams formed at 17 locations, 5 of which remained after the typhoon (Ishizuka et al., 2012). The debris from the Nojiri landslide flowed along a torrent and into the swollen Kumano-gawa, and diverted the flow to the opposite bank, where residential houses were flooded and 8 people were killed or swept away. Landslide debris that rushed into the Kumano-gawa at Uguhara caused a tsunami that destroyed a hydropower station 1 km upstream (Fig. 3). The landslide at Shimizu hit the houses on the opposite side of the Kumano-gawa and 11 people were killed or missing (Figs 2C, 3).

The velocities of landslides are inferred to be on the order of tens of meters per second based on the durations of landslides determined by seismic record. The locations and durations, which ranged from 30 to 100 seconds, of 9 landslides were identified from seismic records (Yamada et al., 2012). The maximum speed of the debris at the Akatani landslide has been precisely estimated to be 80-100 km h^{-1} based on the travel time of 50 seconds from the seismic record (Yamada et al., 2012) and the travel distance of the weight center of debris, 514 ± 33m.

The largest landslide was the Kuridaira landslide, which had an area of 548,500 m^2 and a volume of 14 million m3, and the second largest was the Akatani landslide with an area of 423,700 m^2 and a volume of 8.2 million m^3 (Fig. 2A). The number of landslides decreased with increasing area (Chigira et al., 2012). The relationship between the number of landslides and their area differs from case to case, but interestingly, the parameters for the landslides caused by typhoon Talas and those caused by the 1889 typhoon in the same area (Hirano and Ohmori, 1989), were similar.

The Kii Mountains are predominantly underlain by the Shimanto accretionary complex, which is Cretaceous to Paleogene–Early Miocene in age with minor occurrences of Miocene granite and sedimentary rocks. The complex consists of foliated mudstone, sandstone, acid tuff, chert, and greenstones. It is dominated by broken formation, of which beds are broken and discontinuous, and mixed rocks with a block-in-matrix fabric, and many thrust faults have developed.

The area affected by typhoon Talas was subjected to a very similar typhoon in 1889, which caused 1492 fatalities and many large landslides and floods (Hirano et al., 1989; Makihara, 2012). The disaster in 1889 caused so much damage that 2489 people were forced to move to Hokkaido from Totsukawa village.

2.2 Gravitational slope deformation before catastrophic landslide

All of 14 landslides we surveyed in detail have been found to have occurred on slopes with geomorphic features associated with gravitational deformation. Images of the slopes affected by 10 landslides, derived from DEMs with a resolution of 1 m, clearly showed that all had one or more downslope-facing arcuate scarplets (small scarps) along their future crowns prior to the catastrophic failure (Fig. 4). These scarplets and linear depressions developed as a result of gravitational deformation that preceded the failure, and their scale may represent the strain state of the slopes immediately before the catastrophic failure. The scarplets observed along the slope lines on the topographic profiles varied from 2 to 62 m in length, 2 to 50 m in height, and from 33° to 45° in inclination. They are probably the upward extension of subsurface slip surfaces, but their shape could have been modified by degradation (weathering and failure), making them larger than at their time of formation. Scarplets appeared as an arcuate band or as a zone of arcuate bands. The ratio of the length of a scarplet to the whole slope length (measured along the slope line on a plan view) varied between 5% and 21%, and could be interpreted as an indication of strain levels prior to failure (Moriwaki,

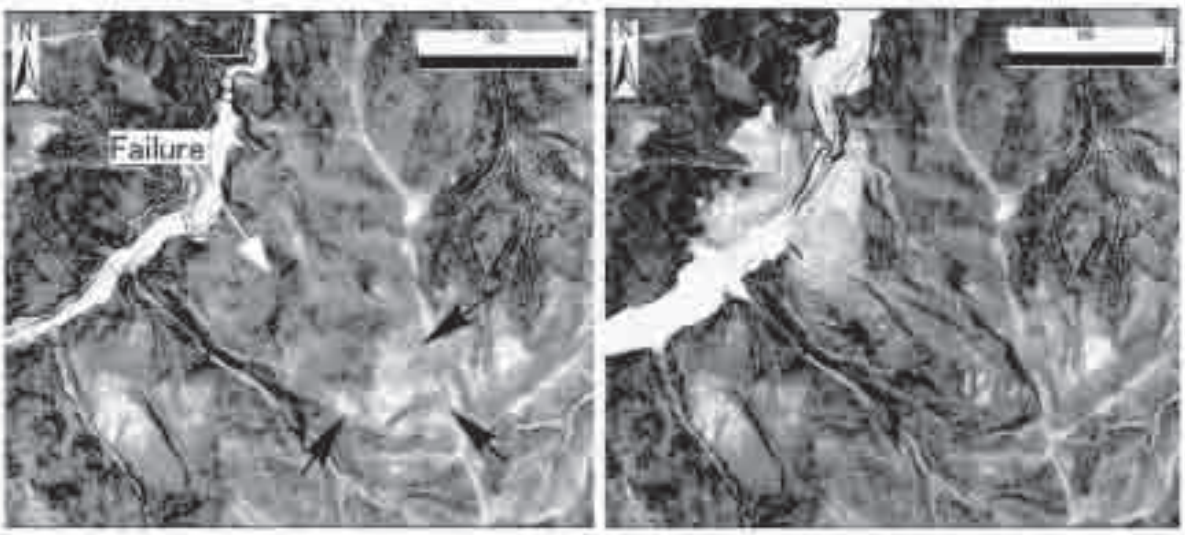


Fig. 4 Slope images before (left) and after (right) of the Akatani landslide. We see dark lines (scarplets) shown by arrows along the future head of the landslide.

2001; Chigira, 2009). The geomorphic strains we obtained for the landslides induced by typhoon Talas may be the first reliable values recorded immediately before a catastrophic failure. Moriwaki (2001) examined strain levels on pre-failure slopes using soil experiments and displacement monitoring in the field, and found that there is a critical strain level that develops before a catastrophic movement occurs. The critical strain could be interpreted as the strain level that develops just before the peak strength of a slope is reached, after which a large drop in stress occurs and the slope fails catastrophically (Sassa, 1985). Petley and Allison (1997) examined the mechanical behavior of London clay using triaxial testing under different confining pressures, and found that it failed in a brittle manner and with a large drop in stress after the creep stage, behavior which they proposed was associated with the growth of micro-cracks. This creep stage and micro-crack growth may correspond to the gravitational deformation of slopes, and consequently to the formation of scarplets and linear depressions as a topographic expression of the deformation.

The landslides we surveyed could be classified according to the nature of the preceding gravitational deformation; i.e., sliding on wedge-shaped discontinuities, buckling on a dip slope, and flexural toppling on a slope with a steeply dipping beds or other foliation. Sliding, which involves discontinuities that form wedges, was the major type of the gravitational deformation in 12 of the 14 landslides surveyed.

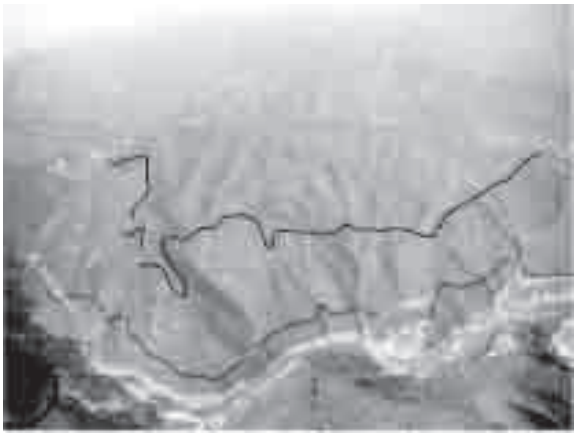


Fig. 5 Landscape of upstream catchment of the Kumano-gawa. Solid and dotted lines are upper and lower slope breaks.

3. Topographic evolution and landslides: Hiraishi and Chigira (2011) found paleosurfaces incised by rejuvenated rivers to create topography of valley in valley (Kelsey, 1988). Inner valley and outer valley is bounded by generally well-defined slope breaks. Two-storied valley in valley, which

accompanies two slope breaks, is widely developed in the upstream of the Kumano-gawa, which is clearly seen in Fig. 5. The higher slope breaks, which are dominant, are up to 250 m high from the present river bed, and the lower up to 100 m. The slopes exposed on the side slope of inner gorges are undercut and destabilized. Dip slopes in particular are destabilized much more than infacing slopes and there occur gravitational slope deformation. Some part of the gravitationally deformed area has a well-defined outline, which separates the sliding mass from the surrounding deformed area. Detailed study areas are upstream tributaries of the Kumano River, Nakahara River and Ten River, but paleosurfaces and valley-in-valleys are understood to be widespread in the Kii Mountains.

The slope development history described above and shown in Fig. 6 gives a background of landslide hazard in this area and could be used as a basis for landslide hazard zonation. Landslide susceptible areas are dissected area in Fig. 6 and undercut slopes above the slope breaks particularly on dip slopes. Actually, the deep-seated catastrophic landslides induced by typhoon Talas occurred on the newly dissected slopes, which suggests that the small-scale gravitational slope deformations described in the previous chapter occurred on slopes thus destabilized by young incision. In addition, most of 1889 rainstorm-induced landslides also plot on the newly incised slopes.



Fig. 6 Schematic sketch showing the evolution of slopes in the Kii Mountains. 1: Old surface; 2: rejuvenation of the river forming the upper slope break and inducing gravitational slope deformation; 3: next rejuvenation of the river forming the lower slope break and further gravitational slope deformation.

4. Conclusions: Typhoon Talas 2011 induced many catastrophic landslides, causing severe damage. Detailed topographic data with 1-m resolution clarified that they had been preceded by preparatory gravitational deformation, which appeared as small scarps and linear depressions. Those gravitational deformations occurred on slopes of inner valleys, which had been made by young river incision in old topographic surfaces. Short time catastrophic phenomena of landslides are thus prepared through geological time.

Acknowledgments

This paper is based on the cooperative research with Matsushi, Y., Tsou Ching-Ying, Yamada M. of the Disaster Prevention, Hiraishi, N. of the Fukada Geological Institute, and Matsuzawa, M. of the Public Works Research Institute. DEM data have been provided were provided by the Ministry of Land, Infrastructure, Transportation and Tourism, Japan, and the Nara Prefectural Government.

Reference

Chigira, M., 2009. September 2005 rain-induced catastrophic rockslides on slopes affected by

deep-seated gravitational deformations, Kyushu, southern Japan. . Engineering Geology, 108, 1-15.
Chigira, M., Matsushi, Y., Tsou, C.-Y., Hiraishi, N., Matsuzawa, M., Matsuura, S., 2012 (in press). Deep-seated Catastrophic Landslides Induced by Typhoon 1112 (Talas). Annuals of Disaster Prevention Research Institute, Kyoto University.
Hiraishi, N., Chigira, M., 2011. Formation of inner gorge and occurrence of gravitational slope deformation in the central Kii Mountains. Transactions, Japanese Geomorphological Union, 32(4), 389-409. (In Japanese with English abstract)
Hirano, M., Ohmori, H., 1989. Magnitude-frequency distribution for rapid mass movements and its geomorphological implications. Transactions of the Japanese Geomorphological Union. 10, 95-111.
Hirano, M., Suwa, H., Ishii, T., Fujita, T., Gocho, Y., 1984. Reexamination of the Totsukawa hazard in August 1889 with special reference to geologic control of large-scale landslides. Bulletin of the Disaster Prevention Research Institute, Kyoto University, 27(1), 369-386.
Ishizuka, T., Okamoto, A., Nakagomi, J., 2012. Emergency structural and non-structural measures against secondary disasters by landslide dams in Kii Peninsula, Japan caused by typhoon Talas in 2011. In Chigira, M. ed. Characterization, prediction, and management of deep-seated catastrophic landslides. Disaster Prevention Research Institute, Kyoto University, 38-44. (In Japanese with English abstract)
Kelsey, H.M., 1988. Formation of inner gorges. Catena, 15, 433-458.
Makihara, Y., 2012. The disaster induced by the 2011 typhoon 12 and the 1889 Totsugawa disaster. Tenki, 59(3), 11-15.
Moriwaki, H., 2001. A risk evaluation of landslides in use of critical surface displacement. Landslides - Journal of the Japan Landslide Society, 38(2): 115-122. (in Japanese with English abstract)
Petley, D.N., Allison, R.J., 1997. The mechanics of deep-seated landslides. Earth Surface Processes and Landforms, 22, 747-758.
Sassa, K., 1985. The geotechnical classification of landslides. Committee for International Exchange of Landslide Technique, ed., Proceedings of the 4th International Conference and Field Workshop on Landslides, Tokyo, 31-40.
Yamada, M., Matsushi, Y., Chigira, M., Mori, J., 2012. Seismic recordings of landslides caused by Typhoon Talas (2011), Japan. GEOPHYSICAL RESEARCH LETTERS, 39, L13301, doi:13310.11029/12012GL052174.

12

Literature and/as tipping point. Reflections on literary form, consciousness and emergence in the eco-literature of the Swiss writer Franz Hohler

Nicholas Saul (Durham)

Abstract

This paper argues that the relevance of system theory for literary studies, unlike other models which attempt to mediate between literature and science, lies in its notion of autopoietic system, which leads to an acceptance of the autonomy of aesthetic cognition in the literary text. A short story of 1982 by the Swiss eco-literature writer Franz Hohler, 'Die Rückeroberung' ('The Reconquest'), serves as an example of the claim. It is argued that this story is structured as a model of the system-environment structure. The story not only attempts to model an tippmg point in the eco-system, but also to foster the emergence of a new eco-consciousness in the observing mind.

In this paper I want to treat the ancient theme of the two cultures, a term coined of course by C.P. Snow in 1959 to connote the dualistic relation of the natural sciences and the humanities, and specifically to treat the relation of science and literature. It will come as no surprise to you to hear that even six decades after Snow we have yet to achieve the goal of constructing the third culture. In fact Snow did not invent the dualism. Attempts to heal the divide go back in cultural history at least as far as the panoramic conspectus of the French encyclopaedists in the mid-eighteenth century, then to synthesising schemes of the German Romantic encyclopaedists around 1800, and then on to the would-be mediating theories of science and culture of Thomas Huxley and Matthew Arnold in the mid-nineteenth century. Attempts to bridge the gap seem in that century to have ended with Wilhelm Dilthey's resigned acceptance in 1883 of the fundamental distinction between the subjective (or inter-subjective) humanities and the objective natural sciences as disciplines respectively of understanding individual internal experience and explaining pure external fact.

Today however we are still hard at it, and there are many more pontificating schemes. Some, such as Foucauldian discourse theory, work at deep-structural level. What we think of as objective facts, this approach would argue, are in truth constellations of data embedded in a so-called discourse, a mode of talking whereby the surface phenomena of speech (representation, communication) are unified and regulated by an implicit and dominant concept or *dispositif*. This deep-structural concept exercises knowledge-power on objects represented, so that we can talk of the discourse on women, on sexuality, on the Orient, and so forth, and which pre-structure our consciousness. We are are invited by Foucault to undermine these pejorative and disadvantaging representations by generating a counter-discourse. Under

Foucault the counter-discourse is very frequently thought of as aesthetic. Aesthetic discourse is to an extent freed from the power of the deep-structural *dispositif* insofar as the work of art does not represent reality, so much as construct an alternative, self-referential world, an autonomous model. That autonomy endows aesthetic discourse with a certain liberating distance, and the consequent ability to reflect on and criticise the prevailing discursive structuring of the world. Now words are not the same thing as discourses in Foucault, but the two dimensions do of course interact. As both non-referential and discursively transparent aesthetic texts have the unique ability to cite and criticise dominant discourses through their master tropes, and it is from this process that counter-discourses emerge. So it is that scientific theories are subjected to aesthetic critique in this way, particularly in terms of their implicit moral and political values, but also, if more rarely, in terms of their scientific validity. Hence counter-discourses can also catalyse the emergence of a new episteme, the global mode of representation in an epoch of history, analogous to T.S. Kuhn's notion of the paradigm.

Another way of mediating the aesthetic and the scientific is the theory of creativity. If reality is thought of as a mental construct, both aesthetic and scientific representations of the world can be thought of as creating reality. There are theories (John A. McCarthy) which link contemporary neurobiology with aesthetic creativity by reducing both in ontological terms to the spontaneous and unpredictable firing at microbiological level of neurons between synapses in order to create new and cognitively more powerful representations of the world. Hence both art and science are generated by the same creative process and generate the same type of cognitive added value in different orders of communication.

The recent advent of Daniel Dennett in philosophy and E.O. Wilson in science has, perhaps more promisingly, made Darwinian anthropology a powerful force for mediating between science and literature. Dennett, famously, insisted that what was true of evolutionary algorithms must be true of animal behaviour, and what was true of animal behaviour must also be true of human behaviour, including aesthetic production and consumption. Thus Darwinism was tantamount to a universal epistemological acid, dissolving and reconstituting all possible non-Darwinian theories of anything as expressions of evolutionary theory. This sounds plausible for human psychology, the object of most art and much evolutionary science. But these approaches are stronger in theory than practice, that is, hermeneutically, in their application to the task of interpreting the text. Stephen Pinker for example sought once to explain the origin and function of art Darwinistically, as 'cheesecake for the mind', as a concentrated, pleasurable dose of cognitive sugar during leisure time from hunting and gathering. Joseph Carroll, perhaps the leading hardcore advocate of literary Darwinism, rejects all scientific and aesthetic constructivism as non-referential, therefore self-indulgent

and self-defeating isolationist solipsism. Rather, the production and reception of aesthetic illusion is a learnt behaviour with a strictly referential purpose, namely, cognitive training of adaptive behaviours; mainly, it must be said, with the function of training us better to play the mating game (which is how he reads Jane Austen). Others, like Brian Boyd, see specific forms of writing, narrative in his case, as a learnt method of focusing and sharing attention socially, which high art develops. Michael Tomasello, similarly, sees narrative as the very foundation of society and the origin of cultural evolution by sharing information with others held to possess intentionality according to the theory of mind. Karl Eibl too sees modes of human language, particularly representation, as the origin of cultural evolution, and driver of its changes.

Now cognitive theories of these kinds are certainly becoming more sophisticated, in that, following criticism of Carroll in particular, they no longer all seek reductively to impose a basic model of tertiary age evolutionary psychology (survival, mating, competition, etc.) on the representations of human behaviour in modern literature. But that is really all that can be said in their favour. What these theories offer seems mainly to consist in historical accounts of the likely evolutionary origin of aesthetic practice according to its general adaptive function. Beyond that palaeontological achievement literary Darwinism at its current stage of development is a disappointment, at least to humanities scholars. It tells us almost nothing about the cognitive function and hermeneutic results of literary form, unless you think that sequential narrative as an ordering and exo-archiving of data exhausts the possibilities of form, satsified seemingly with telling us that proudcing art is something characteristically human and has an adaptive purpose. In Foucauldian (and other) terms it also denies the characteristic autononmy of the literary text, which in modern aesthetics is its defining term. It makes the individual cognitive performance of an aesthetic text subservient to a dominant discourse of science, dissolving aesthetic cognition then in tertiary age Darwinian acid. The good old constructivist approaches – gender, colonialism, poststructuralism, Freudianism, etc., – even jaded as they may make us feel in the twenty-first century, all offer richer possible accounts of the cognitive performance of aesthetic discourse.

And yet there does seem to me to be a further approach which can mediate the world of science with the world of aesthetic experience, perhaps more lucidly, than the theories described so far. I mean system theory, as formulated by the biologists Humberto Maturana and Francisco Varela and adapted in his own unique style by Niklas Luhmann into an all-encompassing social theory. This needs no detailed exposition here. The features which make system theory attractive in this context are its basic concepts: system, environment, structural coupling, perturbation. Especially attractive is system theory's foundational notion

of systemic autonomy, which positively celebrates self-referentiality and clearly recognises the fact of aesthetic autonomy. Attractive too to aestheticians is the notion of an organic or social system as an autopoietic organisation, which in its self-reproductive operations generates the unique kind of effects which are not causally *resultant*. Resultant effects derive from the interaction of the parts of the system, generate a quantitative difference, and are by definition predictable. Autopoietic systems by contrast generate *emergent* effects, effects which manifest themselves quasi-irrationally as something more than the sum of the interaction of the system's constituent parts, generate a qualitative difference, and are by definition unpredictable. This is claimed to be tantamount to adaptation and evolution: the reorganisation and restabilisation of the system under new terms. The concept of emergence, in turn, can be deployed retroactively to describe biological or cultural evolution, notably in Niles Eldredge's and Stephen Jay Gould's theory of punctuated equilibria (1972), and, to some extent, to describe a social analogue, Malcolm Gladwell's famous tipping point as a model of the factors conditioning qualitative social behavioural change which is rapid, radical and unpredictable.

In what follows I want to take as my example of the applicability of this approach to the problem of science and literature a short story by an author very well known in germanophone lands, Franz Hohler, entitled 'Die Rückeroberung' of 1982. Hohler (*1943) is not so well known in anglophone lands. 'Die Rückeroberung' has been translated into several languages (not Japanese so far as I can tell), but was translated into English as the 'The Recapture' only in 2009 – although the title, as we shall see, is more accurately rendered as 'The Reconquest', and I shall refer to it as such. Hohler is many-sided. He is, in addition to being a writer, a brilliant musician and satirical cabarettist, who still performs regularly, also on the television. But he is perhaps best known as a writer, both for adults and children. His major adult works, including 'The Reconquest', which has been republished many times in German, are generally associated with the Green movement and ecological issues, then as now far more prominent in the public sphere in germanophone lands than in anglophone territory. Perhaps his major work in literature is a novel of 1989, *Der neue Berg* (*The New Mountain*), which will explain why his œuvre is often classified as eco-literature. Here a new mountain really does appear in the present day, in this case in the city of Zürich, in the form of a nascent volcano. The point of this well-documented text, which features a great deal of geographic and vulcanological research, is cognitive denial, as the scientific and civil powers that be prove incapable of recognising the symptoms of imminent eco-apocalypse – unusual new cracks in the ground – for what they really are. The volcano is thus a metaphor permitting acerbic satire of smug bourgeois mentalities and the violent nemesis which duly

awaits those who persist in old ways of thinking. Ecological issues do play a part in this novel, but social satire in the conventional sense is to the fore.

'The Reconquest' however foregrounds the eco-literary function. Like *The New Mountain* this story is rooted in humdrum everyday Zürich reality, indeed the beginning and the end are set in what is clearly Franz Hohler's very own study. But effect is gained by deploying extraordinary, indeed fantastic ecological or biological events against this apparently normal horizon of expectation. The story mixes dark humour, such as the glorious eco-nemesis of a Zürich traffic jam under the hooves of a herd of rampant deer (RE 11), with chilly satiric comment, as when the shock of the attritional predation of wolfpacks on Zürich children is contrasted with the traditional supine acceptance of the same number of children's deaths in road accidents in another time (15). Ultimately, then, the story can be classified generically as dystopian contemporary science fiction, rather like, say, John Wyndham's *Day of the Triffids* (1951).

To summarise briefly: Hohler's tale tells of changes in the Zürich eco-system, as observed through a writer's window. Here, as in the later *New Mountain*, cognitive denial is strongly thematised, as one day before the narrator's window there appears perched on the television aerial of the house across the street a very large mountain eagle. It is clearly not an escapee from the zoo, since its wings are not clipped, and obviously a total stranger to the urban habitat. Does he not misrecognise the eagle? our first-person narrator asks himself at the time. The inability to grasp the evidence of his own eyes is compounded by the late arrival of another observer, his wife, who half-persuades him that the creature might have been a large seagull, and so forth. The experience is written off as inexplicable or erroneous until some weeks later the eagle reappears, this time with its mate. They nest, breed, and stay, so that both the cultural and eco-systems must adjust to accommodate the new presence. Small pet mammals are kept inside. The occasional cat is taken, but generally the eagles seem to feed on the rat population of the city, and so are welcomed. But this is a beginning, not an end, and certainly no minor perturbation of the system. Weeks later a gigantic deer or, rather, a herd of gigantic deer is discovered in the city's main park, again apparently having manifested itself out of thin air. All efforts to contain them fail. They are too agile to be shot without endangering citizens. When herded together they are a positive threat. The city police resort to setting up a kind of cowboy squad of horseriders armed with lassoes, to capture individual beasts, and the city assumes the aspect of the Wild West. These creatures too appear to have no provenance. Official science, the zoo biologists and others, are unable to explain the intrusions, which of course represent the vanguard of nature's apocalyptic reconquest of the ecocultural space of the city, and which they therefore show no desire to leave. Only too late

do the inhabitants of Zürich discover that deer are part of the food chain in any stable eco-system, in short, that they bring their own predators with them. Thus large wolfpacks suddenly appear. Unlike the eagles, though, they are wider-ranging in their predation. They also prey on humans, particularly smaller humans. The wolves too prove impossible to remove. This perturbation of the established ecology thus entails a far more vehement systemic response. Children are permitted to move in the open only in parties. Even the children are armed with large knives, and the groups are escorted by citizen soldiers with rifles. Within a year this progression comes to its inevitable fulfilment. The wolves are followed by bears, snakes, and their predators too. In every case science is theoretically and practically powerless to explain or reverse the evolutionary tide. The threadbare layman's hopes of the narrator that the advent of new predators will eradicate the previous plague (as with wolves and deer) are inevitably dashed, as the eco-system of course restabilises after each new incursion. Finally, and much more seriously, comes the advance of the vegetable kingdom. Common or garden ivy suddenly begins to grow with incredible speed to gigantic size, similarly other throttling parasite climbers such as Russian vine, wisteria and clematis. They carpet the ground, cloak even the tallest buildings and congest almost all rail and road systems. Otherwise innocent plants such as coltsfoot and ferns magnify their normal size colossally, compete with and soon throttle full-grown normal trees. Massive doses of herbicide produce no effect, aggressive cutting is equally impotent. Eighteen months into the phenomenon, the city has become almost entirely renaturalised, and actually begins to look like a jungle. Machetes are required to traverse it. Movement is restricted to that of armed parties in local neighbourhoods. As the police become irrelevant, equal and opposite changes occur in social behaviour. Local neighbours display immense new solidarity; but robber bands prey ruthlessly on strangers. Even the climate begins to become more severe than hitherto (RE: 22). As the supply of fossil fuels dries up, wood must be scavenged for heating systems. In winter single rooms of the spacious Zürich apartments are shared by all for warmth. As a change of habitat beckons, an exodus of humans begins. And there the story, at least at the level of content, ends, even if it does not close. (Hohler evidently does not see the need to treat the other life kingdoms, of insects or single-celled organisms, since their onward march is surely implicit.)

Now this clearly, again at the level of content, is in biological evolutionary terms a well-informed piece. It seems plausible to suggest that Hohler will have known something of the Eldredge-Gould thesis of punctuated equilibrium in evolutionary processes, and to suggest that this theory of suddenly-precipitated large-scale change or something very like it is adapted by Hohler here as a figure of poetic justice (nemesis, then) and deployed – obviously

with poetic licence – in the service of an ecological polemic. The message seems to be that there is more to nature than was ever dreamt of in your gradualist evolutionary biology. Nature is an empire that can strike back. For so long tamed and domesticated – and the tale includes flashbacks about the usual timidity of animals on human terrain – , there are evolutionary developmental potentials in nature which, prompted no doubt by some ill-understood feature of humanity's reckless technology-driven self-aggrandisement, may generate an evolutionary saltation and a concomitant counter-attack on humanity's eco-cultural habitat. I cannot say whether it is plausible that so many species should all make evolutionary leaps within an eighteen-month period (the narrated time of the story), although Eldredge and Gould do seem to be making the point that deep-time gradualist evolutionary timescales can be drastically foreshortened by such events as those they describe in that essay, or by the pattern of explosive diversification and sudden contingent extinction of species Gould describes elsewhere in his revision of the interpretation of 'wonderful life' in the Burgess Shale of the Cambrian Age.

But this is to treat only one dimension of the tale's cognitive achievement. For 'The Reconquest' not only describes as it were the objective environmental dimension of the eco-apocalypse as I have reconstructed it, but also the evolving consciousness of the observing writer – object *and* subject, then, as a dual and interlinked or, yes, structurally coupled cognitive interest. And it is this where, as you gather, I see the structural analogy of our tale with Maturana's and Luhmann's system theory. The sharply-delineated poetic structure of subject and object in short models the emergent co-evolution of two systems, human system and complex environment. The writer-narrator thus, on a crudely allegorical biographistic reading, stands for Hohler; but it also stands for the writer and artist in general, and ultimately human consciousness itself *in its systemic relation to the environment*. There is a tradition of this dual cognitive interest in German literature, which, like German idealist philosophy, is famously centred on the exploration of subjectivity, to the exclusion, or, in the case of Fichte, inclusion, of everything else. Examples, which Hohler is clearly referencing, would be for those who might wish to pursue the subject, include E.T.A Hoffmann's 'Des Vetters Eckfenster' ('My Cousin's Corner Window'; 1822) and Wilhelm Raabe's *Chronik der Sperlingsgasse* (*Chronicle of Sparrow Lane*; 1857), both of which feature observation through high Berlin windows as symbols of the narrating subject's perspective on external reality and means for analysing it.

So exactly what co-evolves in the writer's subjective consciousness? For a start, *as* the foregrounding of subjective consciousness it notably frames all the events of the tale, which begins and ends with the observing writer writing at his desk. This foregrounding has two

main functions.

First, it breaks down any safe temporal or interpretative distance between the first-person narrator and the reader. It does this by a play with person and verb tense. One of the things the narrator does in a conventional narrative is deliver authoritative information and judgement. Only this can orientate the reader securely, if that is what the writer wants to do. This is why conventional narration deploys a depersonalised, indeed pseudo-scientific third-person narrator or even merely an anonymous narrative instance, with no recognisable voice or concrete features, so as to suggest objectivity. But Hohler's narrator is the opposite of all that. He is clearly a first-person narrator, that is, his perspective is signalled to be neither universal nor hermeneutically authoritative, but subjective and standpoint-bound, existentially exposed to ignorance and a limited horizon of understanding. We, the reader, see the world through his eyes only, and tend to share that experience in an equally exposed and personal way, with a dramatic thrill of authenticity if not authority. We are witness, not historian; involved, not distant.

Then there is tense. The story's opening lines seem conventional in temporal storytelling terms, in that they are cast in the imperfect tense: 'One day, as I sat at my desk and looked out of the window ...' ('Eines Tages, als ich an meinem Schreibtisch saß und zum Fenster hinausschaute ...'; 7). On any realist convention of storytelling, narrative must as a rule not only be objective and authoritative, but also be couched in a past tense, generally the imperfect or past continuous, for the plausible reason that – if we are to sustain the credibility of the illusion – the narrator needs logically to know the end of the tale before he begins to tell it, and so adopts the form of looking back and remembrance, which offers the firm basis for sound interpretation, affirmation of meaning, existential control. But the past tense is precisely *not* consistently used in 'The Reconquest'. In fact, present-tense narration frames and dominates all the inset past-tense storytelling. At the beginning, already in the second sentence, we have the characteristic shift: 'I have to add that I live in Zurich and that in our country eagles are to be found only in the Alps' ('Ich muß dazu sagen, daß ich in Zürich wohne und daß Adler bei uns nur in den Alpen vorkommen'; 7). The shift into the present is deftly reiterated in the last section: 'Now autumn is on the way and noone knows how we are supposed to keep going' ('Jetzt geht es gegen den Herbst zu, und niemand weiß, wie es weitergehen soll'; 21). From here Hohler keeps the narration in the present until the very end, as the concluding lines reveal: 'and here I sit and wonder whether there is still any point in leaving the city, or whether all this is only the start of something which will spread out unstoppably from this place' ('und ich sitze da und denke darüber nach, ob es jetzt noch einen Sinn hat, die Stadt zu verlassen, oder ob das alles nur der Anfang von etwas ist, das sich von

hier aus uneindämmbar ausbreiten wird'; 23). Thus beginning and end are set in fact in the exact same time and place, at the writer's window in his Zurich apartment, and the events narrated are his attempt, exercising his writer's vocation, to reflect upon and understand the events of the last eighteen months, and to share them in some purposeful and meaningful way with the reader. The problem is, of course, that, he fails. Moreover, whilst the story stops, it does not end (is not, in the literary jargon, *closed*). In fact this narrator is in the same position as, say, the commentator of a live football match. He narrates on the same time-plane as the events narrated, does not know what the end of the story is going to be, attempts to interpret what has happened and to predict what will come, but otherwise lives, with his viewer, in a state of constant suspense – until the end. Then, for the football commentator at least, all is revealed and sense is made of the outcome. Except that there this analogy ends. In the case of 'The Reconquest', our commentator experiences no *result*, for the result is still *emerging*. The apocalyptic events narrated erupt existentially into the frame of the present, as a present danger, then, and ... so it goes on. The writer – and the reader – are thus left hanging on a cliff (like the end of a famous film) or ... perched at a tipping point: stay or go, understand at any price, select for change or not.

This, then, secondly, is why I think this eco-tale is interesting and describable in system-theoretical terms, for here we have co-evolution of two systems, the system of observing consciousness and its complex environment composed of many other systems, *enacted* before our very eyes. We can describe probably all of what goes on in the environment in terms of evolution by punctuated equilibrium (if you will allow Hohler and me some poetic licence). But we can also describe what happens in the writer's consciousness, and which is equally the subject of analysis in this tale, in terms of Luhmannian system theory, as the emergence of something new, namely a new ecological consciousness on the way to itself if not yet arrived; *as* emergent, then. So for example we can see clearly here the difference of a natural and a social system (GLU 1997: 53): a natural system evolves in reponse to internal genetic variations and their relation to the environment; a social system evolves in response to a pertubation or irritation (Maturana 1987: 111; GLU 1997: 187). We can reconstruct the impact of genetic changes beyond the window. But inside the window our focus is on changes of consciousness. That so prominently foregrounded window symbolises evidently the structural coupling of system and environment, their mutual separation and difference and yet interdependence, indeed interpenetration, as it appears to the observing consciousness. The writer's attempts to grasp the meaning of what is happening out there beyond the window correspond to the Maturana-Luhmann notion of environmental perturbations, whereby the system-environment border is respected, but changes in the milieu

entail by structural coupling internal dissonances in the system, which in turn require restabilising responses in the form of recursive feedback loops of thought and adaptive action, as typified by the selective withdrawal of the entire family to the writer's study for the coming winter when oil supplies will have dried up. Or the other measures in the social system beyond the window. Ultimately, the writer's consciousness is dominated by the incipient incompatability of the social system and its milieu, the consequent threat to the viability of the system itself, which precludes a re-entry, and the final threatening inability to reproduce. Here perhaps is the most Luhmannian dimension of the story: the observation, like all observations, happens along a time axis which constitutes before and after, comes itself then too late, after the selective event itself has happened and entailed its changes (GLU 1997: 188). And this is well symbolised by the disjunction between the fictional writer's consciousness and that of the reader.

Just a couple more points which follow: In respect of that disjunction of figured writer and reader, the story also enacts pretty well the Luhmannian notion of art in a social system. In a social system, every operational action needs to be prompted by some kind of rudimentary observation, so that art is already the observation of an observation. In 'The Reconquest' the reader observes just so, as the observer of the observer at that second order of communication (GLU 1997: 126) – and does so critically of course. 'The Reconquest', then, reveals itself in a thoroughly Luhmannian way as a means for society to describe itself to itself, or to observe itself (GLU 1997: 107-108). In offering as eco-art a radical redescription of the relation of system and environment, it can be regarded as a prime example of Luhmannian 'world art' (or better: 'world-view art'), which attempts to make plausible the implausibility of the world it recreates in the illusion, and so to offer an alternative, and meaningful – emergent – future course of action.

Finally, we have here an approach to the big problem I adumbrated at the outset: the mediation of literature and science. Here, an approach frankly defined by science (or Maturana and Luhmann anyway) treats literature in such a way so as to respect both the non-referentiality autonomy of the text (as system theory can hardly fail to do) and its autonomous cognitive performance. In the case of 'The Reconquest' that cognitive power is indeed harnessed to perform a scientific task which, paradoxically, perhaps only art can do: lead consciousness through its unique mode of intuitive self-reflexivity to a ecological-behavioural tipping point, and perhaps nudge it over. Could this be the medium of a third culture?

Bibliography

Franz Hohler (1982): DIe Rückeroberung. Erzählungen. Munich: dtv.

Michael Bauer, Klaus Siblewski (eds.) (1993): Franz Hohler. Texte, Daten, Bilder. Hamburg, Zurich: Luchterhand.

Axel Goodbody (2006): Nature's Revenge: The Ecological Adaptation of Traditional Narratives in Fifty Years of German-Speaking Writing. In: Tamkang Review 37: 1-28.

Jürgen Barkhoff (2007): Franz Hohler: A Magical Realist from Switzerland. In: Sabine Egger, Patrick Studer (eds.): From the Margins to the Centre. Itish Perspectves on Swiss Culture and Literature. Oxford: Peter Lang, pp. 325-343.

Niles Eldredge, Stephen Jay Gould (1972): Punctuated Equilibria: An Alternative to Phyletic Gradualism. In: Thomas J.N. Schopf (ed.): Models in Paleontology. San Francisco: Freeman, Cooper, pp. 82-115, 217-243.

Stephen Jay Gould (1989): Wonderful Life. The Burgess Shale and the Nature of History. London: Hutchinson.

Humberto R. Maturana, Franscisco J. Varela (1987): Der Baum der Erkenntnis. Die biologischen Wurzen der menschklichen Erkenntnis. Munich: Goldmann.

David Blitz (1992): Emergent Evolution. Qualitative Novelty and Levels of Reality. Dordrecht, Boston, London: Kluwer.

Claudio Baraldi, Giancarlo Corsi, Elena Esposito (1997). GLU. Glossar zu Niklas Luhmanns Theorie sozialer Systeme. Frankfurt a.M.: Suhrkamp.

Holger Krapp, Thomas Wägenbaur (eds.) (1997): Komplexität und Selbstorganisation. "Chaos" in den Natur- und Kulturwissenschaften. Munich: Fink.

Thomas Wägenbaur (ed.) (2000): Blinde Emergenz? Interdisziplinäre Beiträge zur Frage sozialer Evolution. Heidelberg: Synchron.

Malcolm Gladwell (2000): The Tipping Point. How Little Things Can Make a Big Difference. NY: Little, Brown.

13

Emergence and reduction: where is the evidence?

Robin Findlay Hendry
Durham University

Summary: Do any real systems display emergent behaviour? The existence of so-called 'weak' emergence is uncontroversial, because it is understood in epistemic terms: emergent behaviour is that which is hard to understand or predict on the basis of the principles governing the parts of a system. Weak emergence is in the eye of the beholder, like faces in the clouds: they're there if you are disposed to see them. The existence of strongly emergent behaviour, on the other hand, is thoroughly controversial: those who reject it do so because it is either (i) incoherent, (ii) incompatible with scientific evidence or (iii) unsupported by scientific evidence. In this paper I argue against claims (i) and (ii), and respond to (iii) by turning the tables: arguments against strong emergence make strong assumptions about the causal closure, or completeness of the physical. I ask what scientific evidence there could be for such principles. The result, it seems, is a stalemate.

1. Emergence in Philosophy The idea of emergence has been debated in many different parts of science and philosophy. In the nineteenth century it was routinely applied, for instance, to chemistry, to life, and to the mind (McLaughlin 1992). More recently it has been applied within condensed matter physics (Anderson 1972). Anderson was drawing on a long tradition in science and philosophy in which emergence is one of a range of positions in a debate over how unified the world is, in respect of the kinds of things it contains, and the laws of nature that govern them. Most philosophers see emergence primarily as a particular kind of response to the mind-body problem: explaining how, given that they seem to be distinct, the mind and the body interact. However, for philosophers of science the emergence of mind is but one aspect of a wider question: the unity (or disunity) of the natural world.

Perhaps because of its wide application, emergence has been characterised in a number of importantly different ways. One familiar approach involves aggregation. A complex system may be emergent if in some sense it transcends its parts, because it has properties that cannot fully be explained by reference to their properties, considered individually or in combination. In such cases, as Philip Anderson has famously put it, '[t]he behavior of large and complex aggregates of elementary particles … is not to be understood in terms of a simple extrapolation of the properties of a few particles' (1972, 393), or in an even more familiar phrase 'the whole is … more than the sum of its parts' (see for instance Davies 2006, x). But it is not clear how this aggregative conception covers all the candidate cases of emergence. The metaphor does not seem to do justice even to physics, where particular causal histories seem to be involved in some cases of emergence (preparation of spin-paired photons in singlet states, for instance). And it is entirely unclear how it should apply in the philosophy of mind. What is it that is aggregated to generate the bearer of a mental property? Clearly, mental properties are not borne by *mere* aggregates of neurons: brains can support mental properties only because they involve dynamically interacting *systems* of neurons that have particular kinds of causal history. But it might be that mental properties are borne by *minds*, which are distinct from the brains that support them. In that case the aggregation metaphor cannot even get going.

Supervenience is another idea that has been closely associated with emergence, at least in twentieth-century philosophy of mind. Supervenience is a relationship between groups of properties: one group, A supervenes on another, B just in case there cannot be variation in respect of A without variation in respect of B. 'British Emergentists' like C.D. Broad (1925) used supervenience to distinguish their position on the mind-body problem from substance dualism: supervenience would tie mental properties to their base physical properties more closely than substance dualism requires, but would allow more autonomy than strict versions of materialism would allow (see below for a discussion of materialism). In fact Jaegwon Kim (2006, 192-3) goes so far as to argue that supervenience of A on B is a necessary condition of A's being emergent from B: anything else just isn't emergence. But recognisably emergentist positions have been formulated, directed at cases within both physics and the philosophy of mind, in which supervenience fails (see Humphreys 1997; O'Connor 2000). Since supervenience is a commitment also of some *reductive* materialist positions (like Kim's own version of type-identity physicalism), it is neither a necessary nor a sufficient part of emergence. Moreover, as Kim himself has been arguing since the 1990s, supervenience amounts only to necessary covariation, so it is of little interest in itself, just as statistical correlation is of much less interest than the causal relationships that underlie it. Of much more interest than supervenience are the dependence relations that might explain it, for it is about the strength of these dependence relations, and whether they leave room for the distinctness of the dependent properties, that emergentism and reductive physicalism disagree. For this very reason it is odd to focus on supervenience.

Discussions of emergence have also associated it with complexity, with feedback mechanisms, with non-linearity, and with downward causation. It is hard to see how any of these (except the last, as we shall see) is conceptually related to emergence, although they may underlie particular cases of emergence. To capture the many different applications it would help to approach the topic at a more abstract level. Emergence, it is sometimes remarked, is a complex relationship between two kinds of thing or property: complex in that it involves a balance between *two* relations (see for instance O'Connor and Wong 2012).

The first of these relations is dependence: roughly, that A arises from, or is grounded in, B. At the very least, dependence implies that A couldn't exist (or couldn't have existed) without B. Dependence comes in different varieties, depending on how the word 'couldn't' is interpreted. *Causal* dependence is where cases of A are *in fact* maintained by cases of B: even supposing that in fact consciousness is only maintained by living brains, this does not imply that it *must* be. In other possible worlds it might be maintained by something else. *Ontological* dependence is a tighter connection between A and B. Some philosophers argue that to be gold (the stuff that is kept in ingots in Fort Knox) just *is* to be composed of atoms that have a nuclear charge of 79: that's what makes something gold. This essentialist claim entails that gold couldn't exist unless atoms with a nuclear charge of 79 existed: a world without atoms with a nuclear of 79 (or indeed a world without atoms) would *necessarily* be a world without gold. Other candidate cases of ontological dependence are provided by the relationships between the properties of wholes and parts, between determinable properties and their determinates, and between dispositions and their categorical bases.

The second relation is novelty. To say that A is causally maintained by B does not entail that A and B are the same property. Nor does saying that having B is what makes something A. Even where these dependencies hold there is room for A and B to be distinct: for A to be something in addition to B, to be *novel* in some way with respect to B. That, really, is the central emergentist thought. Just like dependence, however, novelty comes in a spectrum of strengths, and these strengths account for the

different strengths of emergence. At the weak end of the spectrum of novelty, A and B might only *seem* different, perhaps because we are acquainted with them via different routes, or conceptualise them differently. Perhaps B in fact maintains A, but we do not have the understanding to see how it does (and perhaps *couldn't* have the understanding). At the more substantial end of novelty comes the idea that A is somehow *objectively* distinct from B. What might 'objectively distinct' mean here?

In the last twenty years or so, many philosophers have come to agree that the identity of a property (what makes it the property that it is) is very closely connected with its causal role. A slogan might be: a property is what it does. In this form the principle goes back to the 1980s (Shoemaker 1984), but its history is much longer: the closely related claim that causal power is a requirement for reality were supported in the 1920s by Samuel Alexander (Kim 1998 calls the principle 'Alexander's dictum') and even by the Eleatic stranger in Plato's *Sophist* (it is sometimes also known as the 'Eleatic principle'). Applying this idea in the current context, A is emergent from B just in case instances of A, although dependent on, or arising from, instances of B, confer distinct causal powers. Because A confers distinct causal powers from B, and causal powers are objective features of the world, A and B must be objectively distinct. It is highly controversial whether this last idea of strong emergence applies to any real system (I will address the issue more fully in the next section), because it conflicts with a presupposition of many philosophers and scientists: that the only real causal powers are those conferred by fundamental physical properties. A very closely related idea is that worldly events and processes are governed by just a few basic properties, linked through just a few basic laws. These laws determine the passage of events, to the extent that they are determined. (Note that, if the standard interpretations of quantum mechanics are correct, and we live in an indeterministic world, we can say instead that only the *chances* of events are determined.) If such a principle is true, emergent properties could *not*, as strong emergence requires, confer causal powers over and above those conferred by their physical bases. This principle has been formulated in two slightly different ways: as the principle of the causal closure of the physical, and as the completeness of physics. These are sometimes used interchangeably, but there are potentially important differences. One formulation mentions causation, the other does not. One involves the specific scientific discipline of physics, the other a class of properties, entities or events, which may or may not be picked out by reference to the discipline of physics. For the purposes of this abstract it is convenient to use a single abbreviation (CCP) to name this kind of principle.

Just about every aspect of CCP is controversial. Some philosophers think it cannot be coherently formulated. Others think that, even if it can be coherently formulated, it is plausible only given an outmoded view of the metaphysics of causation. Some think it is so obviously true that it hardly needs arguing for: in fact, the burden of evidence is on those who reject it to prove it false. Others think that is a strong, and most probably false, claim. Some think it is supported by modern physics, and its explanatory power with respect to the discoveries of 'special' sciences like chemistry and biology. Others think it is violated by modern physical theories like quantum mechanics, or that, although physical theories like quantum mechanics have deepened the explanations provided by special sciences like chemistry and biology, the detail of these explanations provides no evidence for CCP.

Whether or not there *are* cases of strong emergence presupposes that there *can be* such cases. Hence any discussion of strong emergence, under the causal understanding set out above, requires a thorough examination of CCP: its proper formulation, its consequences, what evidence there *might be* for (or against) it and what evidence there *is* for (or against) it.

2. Is strong emergence a real phenomenon?

Are any real properties or systems strongly emergent? Whether in the mental or in the (broadly) physical realm, that issue has always been associated closely with the question of the unity of science: how many different kinds of thing does science study? Are the more complex ones anything over and above their more basic components? Does nature express just a few basic laws, which directly govern the most fundamental parts of nature, and through them, the more complex systems they compose? Or does nature display a complex and disunified patchwork of laws (Cartwright 1999)? The classical conception of the unity of science was provided by Oppenheim and Putnam (1958), who made explicit a widespread view of the sciences as hierarchically structured, and predicted their progressive explanatory unification, with the entities of sciences higher up being shown to be complexes constructed out of the entities of sciences lower down. Nagel (1979) provided a formal model of this explanatory unification: the laws of the reducing theory explain (for which read: deductively entail) the laws of the reduced theory. However, no examples of intertheoretic reduction have ever been accepted universally: even the classic case of temperature and mean molecular kinetic energy has always faced objections (see Sklar 1993 for a survey of the issue, and Needham 2009 for more recent criticisms of this case as an example of reduction). Yet temperamental reductionists have always been untroubled by these objections, and temperature is still widely claimed to be reducible to mean kinetic energy (see for instance Loewer 2001 and Papineau 2010). One common response to the failure of classical Nagelian reductionism is that, despite the inapplicability of Nagel's derivational model, classical thermodynamics, chemistry, life and the mind are all 'reducible in principle' to fundamental physics: it is just that the Nagelian derivations are blocked by the sheer complexity of special-science systems, the mathematical intractability of the equations that describe them, or conceptual mismatch between the physics and the special sciences. It was the letter of Nagel's utopian model that was at fault, rather than the spirit of reductionism more generally. Yet there remains widespread (though not universal) agreement across philosophy of mind and philosophy of science that physical properties form a special class whose members are asymmetrically related to the kinds of properties studied by other sciences like chemistry and biology.

If Nagelian reductions are not to be expected, what exactly do the reductionist and the emergentist disagree about? The reductionist might limit their claim to 'reducibility in principle,' but this is far too vague a notion for proper debate (Crane and Mellor 1990), and in recent years the retreat has been to ontological reducibility instead. The broad idea of ontological reducibility is that A is, in some objective sense, 'nothing but' B. Given the discussion in the last section, the obvious way to cash this idea out is in causal terms: A is nothing but B if it confers no additional causal powers. This makes (strong) emergence just ontological *non*-reducibility. The disagreement then comes to centre on CCP, acceptance of which seems to be a litmus test for a thoroughgoing physicalist reductionism. Thus McLaughlin (1992) and Kim (1998) object to versions of physicalism involving only supervenience between the mental and the physical, because they fail to rule out downward causation, as (they argue) any physicalism worth the name ought. One way of putting the problem is as follows: suppose that some group, A, of special-science properties, supervenes on some group of physical properties. If there are law-like connections among the A-properties (special-science laws), these will be reflected in relations among the physical properties, although those relations may well be messy and disjunctive. Given the supervenience, any A-changes will be accompanied by physical changes: should we regard the A-changes as making the physical changes happen, or vice versa? If the former, then there is a sense in which the physical properties are being 'pushed around' by the A-properties (hence there is

'downward causation' from the A-properties to the physical properties on which they supervene). Only CCP rules out this possibility, and there are two striking things about its role in the debate. Not only does CCP justify why the ontological priority of fundamental physics is so widely assumed, but it is also, according to its defenders, the kind of principle for whose support we can appeal directly to discoveries and explanations within science itself.

Few philosophers of mind or science have thought it necessary to make explicit just where in science the evidence for physicalism should be sought, and even fewer have made a detailed case that the required evidence can actually be found there. Notable exceptions to this reticence are Brian McLaughlin (1992), Barry Loewer (2001), David Papineau (2002) and Andrew Melnyk (2004). In a generally sympathetic (and widely cited) discussion of 'British emergentists' like Samuel Alexander and C.D. Broad, McLaughlin argues that their position, although coherent, lacks scientific support. If British emergentism were true then there would exist real systems that are moved by so-called 'configurational' or emergent forces, as opposed to 'resultant' forces which are built up from fundamental physical forces. But as he puts it, there is 'not a scintilla of evidence' that they do exist. It is a common assumption of physicalist arguments for CCP that the burden of proof lies with the emergentist (Loewer and Melnyk argue in similar ways). On the other hand, Nancy Cartwright (1983), has long argued that the evidential burden should be on the other side: the conclusion that some theory is applicable to some phenomenon requires a detailed predictive model of that phenomenon within that theory, rather than an argument-in-principle that covers a whole domain of phenomena, yet is based on generalization from just a few simple cases which may be particularly simple. Thus, for instance, Cartwright would urge caution in moving to the conclusion that quantum mechanics alone is sufficient to explain the whole of chemistry, if this is based only on an elegant treatment of the hydrogen atom (a special case on account of its symmetry), or the hydrogen molecule (also a special case, like other diatomic molecules). The debate on emergence then reaches an impasse, because the two sides cannot agree on what evidence is relevant to it. If the debate on emergence is to move on, then these underlying assumptions about evidential relevance need to be identified and subjected to critical examination.

One reason why reductionists think that the burden of proof should be on the emergentist side is as follows: I will call it the 'Nobel Prize' argument, following a common rhetorical device. Physics is special in one particular way: everything non-abstract—everything that is located in space and time; able to be involved in causal processes—has physical properties. (Abstract objects like numbers and concepts may lack them, of course.) So physical laws apply to everything (non-abstract). Thus physical laws, which govern physical properties, are universal: they *apply* to everything, unlike chemical laws, biological laws etc., which have clear domains of non-applicability. The punch-line is as follows: emergentists think that physical laws are non-universal, and therefore that they have exceptions. But showing that that is true would be a major scientific achievement, of the kind that could earn one a Nobel Prize in Physics. Given that Nobel Prizes are not being given out to emergentists, the empirical evidence for their position must be lacking.

The trouble with this argument is that it confuses two principles of universality. Emergentism is clearly incompatible with one (very strong) principle of the universality of physical law: that physical principles completely determine the behaviour of the systems to which they apply. It is, however, compatible with an alternative principle of universality, the ubiquity of physics. Under the ubiquity of physics, physical principles constrain the motions of particular systems though they may not fully determine them. Some universal physical principles are naturally understood this way, even by

physicists: the second law of thermodynamics, and the conservation of energy are obvious examples. Taken individually, the various force laws which dictate the form of the potential terms in quantum-mechanical Hamiltonians seem also be understood this way, since they operate together to produce the potential term which governs the overall motion of a system. For instance Coulomb's law is not violated in systems in which (say) gravitational forces also act. Robert Bishop (2006, Section 2) makes a closely related point, arguing that 'Physics itself does not imply its own causal closure' (2006, 45). In short, one may accept that physical principles apply universally without accepting that they fully determine the motions of the systems they govern. The universal applicability of physical principles does of course imply that, *when acting on their own*, they fully determine the motions of any system they govern. But this leaves open what happens when other forces are at work.

The difference between the two principles is subtle and easily missed: in defending 'fundamentalism' against Cartwright's arguments against it (Cartwright 1999), Carl Hoefer (2003) rightly distinguishes fundamentalism, the claim that there are 'universal fundamental laws with which all phenomena are in accord' (2003, 1403) from the stronger 'thesis of the reducibility of biology, chemistry, or meteorology to physics' (2003, 1408). As formulated, Hoefer's fundamentalism seems to express ubiquity very nicely: he doesn't explicitly distinguish intertheoretic and ontological reduction, though he clearly intends to deny only intertheoretic reducibility. Therein lies the problem. He commits fundamentalism to the causal completeness of physics (and therefore ontological reducibility) although he defends only ubiquity against Cartwright's arguments. He endorses a quote from Richard Feynman clearly expressing ontological reductionism, but glosses it as ubiquity:

> What the fundamentalist believes in is a sort of no-conflicts thesis, between physical laws and higher-level phenomena. Feynman … expresses it nicely: "For example, life itself is supposedly understandable in principle from the movements of atoms, and those atoms are made out of neutrons, protons, and electrons. I must immediately say that when we state that we understand it in principle, we only mean that we think that, if we could figure everything out, we would find that there is nothing new in physics which needs to be discovered in order to understand the phenomena of life." (Hoefer 2003 note 3, 1408-9, quoting Feynman 1965, 151)

Emergentism is compatible with ubiquity, though not with completeness. Completeness implies reducibility, as we have seen, while the ubiquity of physics does not: this is not surprising, as completeness is a logically stronger principle of the 'special status' of physics than ubiquity. The mere applicability of physical principles to the other sciences requires only ubiquity, and does not rule out downward causation. Emergentism implies no violation of physical laws. In the argument between Cartwright and Hoefer's fundamentalist there are now two versions of fundamentalism in play: one is committed to completeness, the other only to ubiquity. Hoefer's arguments defend only the latter against Cartwright. There are two corresponding versions of Cartwright's anti-reductionist metaphysical picture of a 'patchwork of laws' (Cartwright 1999): one denying only completeness, the other denying ubiquity too. The former is quite close to Broad's version of emergentism, which elsewhere (Hendry 2006, 2010) I have developed so as to involve a *counternomic* criterion for emergence: a property is emergent if it is governed by laws which are different from those it would be governed by if reductionism were true, and the laws governing fundamental physical properties completely determined how it behaved.

To conclude this section, if the debate is to move on, then the arguments for CCP must be identified and subjected to critical examination, with one of the following results: (i) the arguments are found to be broadly correct; (ii) they are improved; (iii) CCP is amended to match the (lack of)

evidence for it; (iv) commitment to CCP is shown to be unjustified. Given that explanatory relationships between physics and the other sciences must bear on CCP, then assessing CCP is (at least in part) work for the philosophy of science. One issue will be whether the unique causal or explanatory role accorded to physical properties is best understood as involving CCP, or the logically weaker conjunction of ubiquity (see above), and the strictness of physical laws governing those properties (see Hendry 2010). Cases from two areas of science—chemistry and condensed matter physics—are uniquely salient to this discussion. Firstly, detailed studies of reduction, emergence and intertheory relations have recently become available for these cases (see for instance Sklar 1993, Batterman 2002, Hendry 2006, 2010, forthcoming). Secondly, explanations in these domains are expressed in the mathematical language of fundamental physics, and apply its theories, such as quantum mechanics. Surely in these domains, if anywhere, the ideas of emergence and reduction can be given precise expression and tested. Surely here, if anywhere, the emergentist side must provide positive examples, described in the language of physics. And surely here, if anywhere, the onus is on the reductionist side to provide recognizably reductive explanations, or to explain why, if they cannot be provided, reductionism is still tenable.

3. Is strong emergence coherent?

The critics of strong emergence we saw in the last section typically accepted that the idea makes sense, but argued that there is no evidence that it has any real examples. However, in a widely cited paper Kim (1999) has argued that emergence is not described in sufficient detail to distinguish between two alternatives, one of which is uninteresting, the other incoherent. In this section I will explain and address Kim's objections.

It is a long-standing worry from temperamental reductionists that the apparent plausibility and interest of emergence depends on its being presented in imprecise terms. In particular, criteria for emergence may fail to distinguish between causal processes in complex systems which are uninteresting because unproblematic for reductionism (whence the apparent plausibility), and cases which appear to be of genuine interest but on closer scrutiny are not, however, coherent possibilities.

Kim rightly points out that the interest and distinctiveness of emergentism as a form of materialism derives from downward causation, understood as higher-level entities and properties having a 'causal influence on the flow of events at the lower levels, levels from which they emerge' (1999, 24). For Kim, the question is how they can coherently be supposed to do this. Following Kim, let *W* be a whole with some property *M*, which is realized by *W*'s microstructural properties, and yet is emergent with respect to them. First distinguish reflexive from non-reflexive downward causation: in the former, 'the whole *W* is a cause of, or has a causal influence on, the events involving its *own* microconstituents' (1999, 26). Then distinguish synchronic and diachronic downward causation, depending on whether or not *W*'s emergent property helps to bring about its own realisers (1999, 28). These two distinctions result in a four-way division of downward causation into synchronic reflexive, diachronic reflexive, synchronic non-reflexive, and diachronic non-reflexive cases. Kim argues that only one category, synchronic reflexive downward causation, is interesting. Non-reflexive downward causation, Kim argues, is 'mundane' (1999, 26), encompassing 'legion' cases like that of a vase falling and bringing about all sorts of changes to nearby molecules as it cuts a swathe through the air and smashes on the ground (1999, 25-26). Emergentists, he rightly points out, would not count this as an interesting case of downward causation even though neither the causally active property of the vase, its mass, nor the causal powers it confers, are possessed by any of its component microparticles. This

is because the mass of the whole vase, and the causal powers it confers, arise in perfectly explicable ways from properties of its component microparticles, hence both are 'resultant' (1999, 26). Diachronic downward causation is similarly ubiquitous and uninteresting:

> I fall from the ladder and break my arm. I walk to the kitchen for a drink of water and ten seconds later, all my limbs and organs have displaced from my study to the kitchen. … It doesn't seem to me that these cases present us with any special mysteries rooted in self-reflexivity, or that they show emergent causation to be something special and unique. (1999, 30)

I think that, for the purposes of judging whether they are compatible with reductive physicalism, these cases of downward causation have been under-described, but we can get on to that in a moment. First note that Kim considers only one category of downward causation to be interesting, and troubling for reductionism: synchronic reflexive downward causation. This is just the category that must be empty, however, given what Kim calls the 'causal-power actuality principle' (1999, 29) according to which, in order to exercise some causal power, an object must *at that time* possess the property which confers it. The causal-power actuality principle, Kim argues, renders it incoherent to suppose that a higher-level property could be part of the process which brings about its realization base. Diachronic reflexive downward causation is not troubling because it allows for a time delay between an object's coming to have the emergent property and its exercising the causal powers conferred by that property.

The upshot of Kim's argument is worth quoting in full:

> We must conclude then that of the two types of reflexive downward causation, the diachronic variety poses no special problems but perhaps for that reason [is] rather unremarkable as a type of causation, but that the synchronic kind *is* problematic and it is doubtful that it can be given a coherent sense. This may be due to its violation of what I called the causal-power actuality principle, but apart from any recondite metaphysical principle that might be involved, one cannot escape the uneasy feeling that there is something circular and incoherent about this variety of downward causation. (1999, 30-31)

Now the causal-power actuality principle looks plausible, and whether or not it fully captures Kim's 'uneasy feeling' about synchronic reflexive downward causation, I think that the emergentist could accept it as a constraint on their understanding of downward causation without in any way undermining the metaphysical interest of downward causation, or its incompatibility with reductive physicalism. Here's why: Kim's four-way categorization of downward causation cuts across what is important in judging whether a particular kind of case is interesting or mundane. In particular, that categorization does not consider the structure of the laws by which a property confers the causal powers it does. In Section 2, however, we saw that the crucial disagreement between the emergentist and the reductive physicalist concerns their differing views of how just such laws are structured. Hence Kim is premature in dismissing cases of non-reflexive downward causation and diachronic downward causation as mundane or unremarkable. Consider the vase: that is an uninteresting case of downward causation just in case the causal powers conferred by a composite body's having a particular mass are fully explicable in terms of properties of its microconstituents and the laws governing their causal powers. To be fair to Kim, there is probably nothing emergent in thye case of the falling vase. The causal processes involving its mass would fail the counternomic test of Section 2: they would be no different were they determined by the more basic laws governing the stuff of which the vase is made. But that doesn't mean there are no interesting candidates for emergence: elsewhere (2010) I have identified molecular structure as one case where the counternomic criterion could be argued to apply. The argument turns on the symmetry properties of quantum-mechanical Hamiltonians

for isolated molecules, which are difficult to reconcile with the lower symmetries of chemically interesting molecules (see Sutcliffe and Woolley 2012). What breaks the symmetry of the isolated molecule? One would appeal to the environment, but that environment is made up of molecules to which the same argument applies. What breaks *its* symmetry? In order to understand the origin of molecular structure, we must appeal to a composite system (molecule plus environment) which is constraining the behaviour of one of its parts (the molecule): a configurational force. The emergentist will then point out (i) that molecular asymmetries (such as, for instance, the asymmetrical charge distribution of a molecule of hydrogen chloride) are what ground the causal powers of chemical substances, and that (ii) such asymmetries can plausibly be regarded as maintained by configurational forces. Perhaps emergence is coherent and interesting after all.

References

Anderson, P.W. 1972 'More is different' *Science* 177, 393-396

Batterman, Robert 2002 *The Devil in the Details* (New York: Oxford University Press)

Bishop, Robert 2006 'The hidden premise in the causal argument for physicalism' *Analysis* 66, 44-52

Broad, C.D. 1925 *The Mind and Its Place in Nature* (London: Routledge & Kegan Paul)

Cartwright, Nancy 1983 *How the Laws of Physics Lie* (Oxford: Clarendon Press)

Cartwright, Nancy 1999 *The Dappled World: A Study in Complexity* (Cambridge: Cambridge University Press)

Crane, Tim and D.H. Mellor, 1990 'There is no question of physicalism' *Mind* 99 185-206

Davies, Paul 2006 Preface to Philip Clayton and Paul Davies (eds.) *The Re-Emergence of Emergence* (Oxford: Oxford University Press)

Oppenheim Paul and H. Putnam 1958 'Unity of science as a working hypothesis' in H. Feigl, M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science* Volume II (Minneapolis: University of Minnesota Press)

Feynman, Richard 1965, *The Character of Physical Law* (Cambridge MA: MIT Press)

Hendry, Robin Findlay 2006 'Is there downward causation in chemistry?' In Davis Baird, Lee McIntyre and Eric Scerri (eds) *Philosophy of Chemistry: Synthesis of a New Discipline*. (Dordrecht: Springer)

Hendry 2010 'Ontological reduction and molecular structure' *Studies in History and Philosophy of Modern Physics* 41, 183–191

Hendry, Robin Findlay (forthcoming) *The Metaphysics of Chemistry* (Oxford: Oxford University Press)

Hoefer, Carl 2003 'For fundamentalism' *Philosophy of Science* 70, 1401-1412

Humphreys, Paul 1997 'Emergence, not supervenience' *Philosophy of Science* 64, S337- S345

Kim, Jaegwon 1998 *Mind in a Physical World* (Cambridge, MA: MIT Press)

Kim, Jaegwon, 1999, 'Making sense of emergence' *Philosophical Studies* 95, 3-36

Kim, Jaegwon 2006 'Being realistic about emergence' in Philip Clayton and Paul Davies (eds.) *The Re-Emergence of Emergence* (Oxford: Oxford University Press)

Loewer, Barry 2001 'From physics to physicalism' in C. Gillett and B. Loewer (eds.) *Physicalism and its Discontents* (Cambridge: Cambridge University Press)

McLaughlin, Brian 1992 'The rise and fall of British emergentism' in A. Beckermann, H. Flohr, and J. Kim (eds.) *Emergence or Reduction?* (Berlin: de Gruyter)

Melnyk, Andrew 2004 *A Physicalist Manifesto* (Cambridge: Cambridge University Press)

Nagel, Ernest 1979 *The Structure of Science* Second Edition (Indianapolis: Hackett)

Needham, Paul 2009 'Reduction and emergence: a critique of Kim' *Philosophical Studies* 146, 93-116.

O'Connor, Timothy 2000 *Persons and Causes* (Oxford: Oxford University Press)

O'Connor, Timothy and Hong Yu Wong, 2012 'Emergent properties' in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, (URL = <http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/>)

Papineau, David 2002 *Thinking about Consciousness* (Oxford: Clarendon Press)

Papineau, David 2010 'Can any sciences be special?' in Cynthia Macdonald and Graham Macdonald (eds.) *Emergence in Mind* (Oxford: Oxford University Press)

Shoemaker, S. 1984 'Causality and properties' in his *Identity, Cause, and Mind* (Cambridge: Cambridge University Press)

Sklar, Lawrence 1993 *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics* (Cambridge: Cambridge University Press)

Sutcliffe, Brian T. and R.G. Woolley 2012 'Atoms and molecules in classical chemistry and quantum mechanics' in Robin Hendry, Paul Needham and Andrea I. Woody (eds.) *Handbook of the Philosophy of Science*, Volume 6: Philosophy of Chemistry (Amsterdam: Elsevier)

14

The Dynamics of Culture Change: Gradual or Punctuated

Seth D Kunin
Durham University

In this paper we examine three aspects of structural variability and transformation. The first two parts focuses on internal and boundary conditions of variability. They indicate that while variability is found in both, with boundary variability being particularly strong, in both cases essential structural integrity is maintained. The third section, focusing on structural transformation suggests that at least in the two cases discussed significant structural transformation appears to be catastrophic rather than gradual and require some form of external catalyst. This paper, however, leaves a significant question for future research, that is, is there also a form of significant structural transformation that is gradual and evolutionary. While it is clear that all cultures transform over time, it is not clear that this is associated with significant structural transformation.

The dynamics of culture change raise important questions in relation to the nature of underlying structure. While surface or narrative level changes in culture are both constant and observable, transformations in underlying structure appear to be both significantly slower and more difficult to pin-point. While the interrelation between these two levels can be easily confused, as in Levi-Strauss' distinction between hot and cold societies; transformation on the narrative level, even significant and rapid transformation may not signal transformation at the level of underlying structure. Nonetheless, given that cultures have different underlying structural configurations, and that even closely related cultures or traditions, as for example the Israelite and early Christian communities, have different underlying structures, it is clear that underlying structure must also have mechanisms for transformation.

Examination of these mechanisms for transformation at both the narrative and structural levels suggests two possible avenues for transformation – catastrophic and evolutionary. The catastrophic form of transformation is easily observable, particularly in relation to the surface levels. This is where there is an observable impact of a significant internal or external force, which transforms the observable cultural forms within a very short period of time. The most obvious examples of this type of transformation are a revolution which at least apparently reorders the structure of social relations, or the impact of an external force, for example through invasion or conquest, that again reorders or imposes new forms upon a culture. While the surface level transformations brought about by such events are apparent, it is not as clear whether underlying structure is equally transformed – a revolution may reorder power relations in a society without transforming the structures of thought or categorization. Nonetheless, such events can provide at least one mechanism where by underlying structure is significantly transformed, if for example, the new social relations cannot fit into its previous system of categorization.

Understanding the evolutionary processes of structural transformation requires a rethinking of a number of aspects of culture and underlying structure (both abstractly and in the concrete ethnographic context). The questions relating to culture work on two levels: internal and external. The internal questions challenge the nature of structure as a monolithic system; is structure culturally monolithic, or are there internal divisions and variations in underlying structure – and if there are what is the nature of these variations. The external issues relate primarily to the nature of boundaries between different cultural systems. The key issue relates to permeability. Are cultural

boundaries essentially permeable, and does this permeability extend to underlying structure. A subsidiary question relates to the possibility of different types of permeability and they relation to different underlying structural equations.

While evolutionary models often assume slow gradual change, the dynamics of underlying structural change may require a model closer to punctuated equilibrium – based on internal structural processes rather than external catastrophic change. The distinction here relates primarily to the internal structural process. Variation within structure at the cultural level is often characterized by nuancing at the level of emphasis, rather than significant structural difference – thereby maintaining intra-cultural communicability. Structural transformation, whereby the structural equation is changed rather than differentially emphasized, moves outside the sphere of intra-communicability – this change analogous to that of speciation may require a more significant point of transformation than allowed for in the traditional evolutionary model.

This paper focuses on the three key aspects of this question drawing on empirical ethnographic material. The first section explores the dynamics of internal complexity and its relation to underlying structure. It argues that despite significant narrative level complexity, perturbation of underlying structure remains minimal. The second section explores the dynamics of boundary areas, and the role of underlying structure in shaping the nature and movement within these areas. The argument suggests that differential structures shape the nature of movement of mythemes, ritemes and cultemes as well as being potentially shaped and transformed in the process. The final section examines the nature of transformation of one structure into another, suggesting that such transformations, while emerging from structures that have been deformed through different forms of emphasis and nuancing, are substantially catastrophic rather than gradual. This section, however, is meant to be indicative, as much of the material leading up to the transformation examined remains to be analysed.

Section 1: Dynamics of Internal Complexity

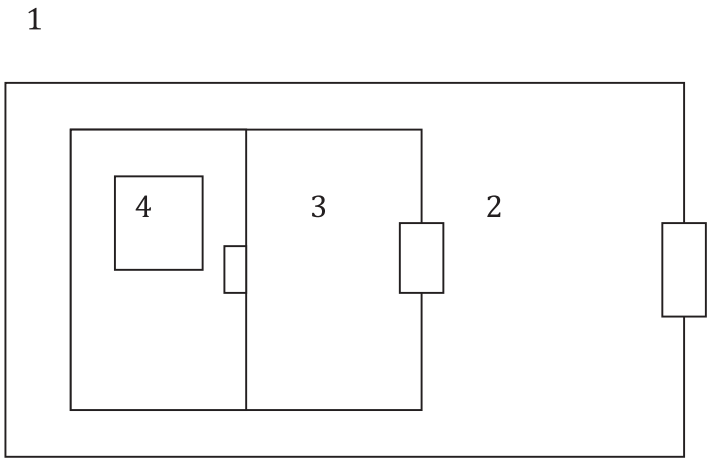
While cultural complexity has been a significant feature of most anthropological discussions, it has hitherto played little or no role in the structuralist analyses of culture. These discussions have tended to focus on monolithic structures, ignoring the possibility of internal variation and structural complexity arising from the complex groupings of individuals within cultural communities. While the monolithic approach was a useful huristic, the analysis of potential complexity is essential as part of an argument about processes of evolution and transformation – if internal complexity as absent, then it external forces become the only transformative agents, while if complexity is present than internal forces can play an equally significant role in the process.

The biblical (Old Testament) text, emerging from Israelite community over a 1000 year period provides an interesting case study of differential uses of structure (by different synchronic and diachronic groups and contexts). While the issues could be mapped on any number of cultural areas, for example sacrifice, or concept of the 'state' The biblical text itself, as a text with a complex historical development, and composed of a range of material created by differing communities and interest groups contains a range of different models of sacred place. Despite these differences in origin, it should be emphasized from the outset, that all of the models have a common underlying structural model. This structural unity is due in part to two mechanisms. First, the text as a whole, in its final form, is the product of redactors and editors, who would have unconsciously reshaped the underlying structure of the material they used, to fit into the underlying

structure of their own time and cultural context. Second, as the communities and interest groups that created the different models were (arguably) part of a wider Israelite cultural context, they would have largely shared the underlying structure of that wider community, and thus even while having some difference in nuance, the structures they utilized would be closely related to those of Israelite culture.

The text, however, contains a dominant model, with variations. This model, the **centralized**, is exemplified by two variations: the **dynamic variant** as depicted initially in Exodus and refined in the later books of the *Torah*, that is, Leviticus and Numbers; the **static variant** is found in the Temple of Solomon described initially in 1 Kings (though an earlier form may be found in the Temple in Shiloh described in Samuel). The dynamic and static variants share a set of abstract principles that justify categorizing them as a single model; it is also likely that the dynamic variant is reading back the static variant into earlier Israelite myth/history as a means of justifying the role and significance of the Temple in Jerusalem. Our discussion will primarily focus on the dynamic model, as it contains an abstraction of the significant features, and therefore allows the clearest exemplification of the underlying structures of the centralized model, whether dynamic or static, as a whole.

One of the most detailed versions of the dynamic variant is found in Exodus 25-27. It is significant for our discussion that this text is followed by the consecration of Aaron and his sons as Priests. This is important as one of the key features of the use of place, it the separation of spaces of differential access – with the Priests and the High Priest having sole access to the most significant areas of sacred place.



1: The camp
2.:The Court
3: The Holy
4: The Holy of Holies

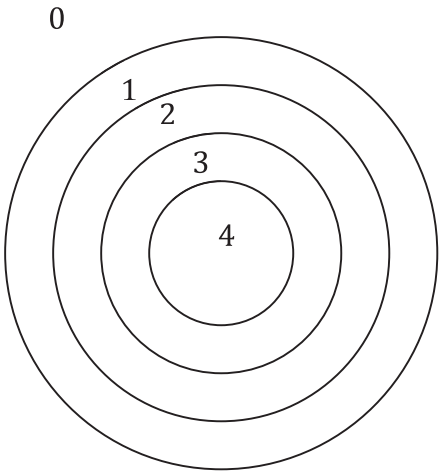
Figure 1: Dynamic and Static Centralized Sacred Place

Dynamic sacred place (Figure One) was divided into at least four zones. Moving from less sacred to more sacred these included: the Camp (that is, the moving space in which the Israelites set up their encampment in the desert during the forty years wandering between the exodus and the conquest of the land), the Court, the Holy and the Holy of Holies. It is also important to include the Camp as a whole as part of this model of

sacred place. The Camp, like the other elements of sacred place was defined both in respect of inclusion and exclusion and by complex rules of purity.

The different zones are demarcated in two primary ways. First is the interrelationship between the human and the divine. This demarcation relates to differential access and presence of the divine. The camp is characterized by access to all Israel, moving westward into the Court access is limited to the Levites, the Holy is limited to the Priests (Cohanim) and the Holy of Holies to the High Priest. The divine presence moves in the opposite direction – it is most present in the Holy of Holies and least present in the Camp – and by implication either absent or minimally present in the world.

Purity is the key second form of demarcation; it is both a defining feature of the different zones of holiness and a quality (or lack thereof either temporarily or permanently) of the human (and non-human) actors within the different sacred zones. Purity as a definitional feature is directly related to differential access and presence of the divine: purity is most strongly associated with the Holy of Holies and least associated with the Camp, the world outside of the Camp is the location if impurity. In relation to human purity, only those individuals of the highest purity are allowed to move into the locations of greater sanctity. Within the non-human actors, for example sacrificial animals there is a similar winnowing process with only the most perfect and specific being permitted for sacrifice.



0: The World/Nations
1: The Camp/Israel
2: The Court/Levites
3: The Holy/Cohanim
4: The Holy of Holies/The High Priest

Figure 2: A Concentric model of Israelite Centralized Sacred Place

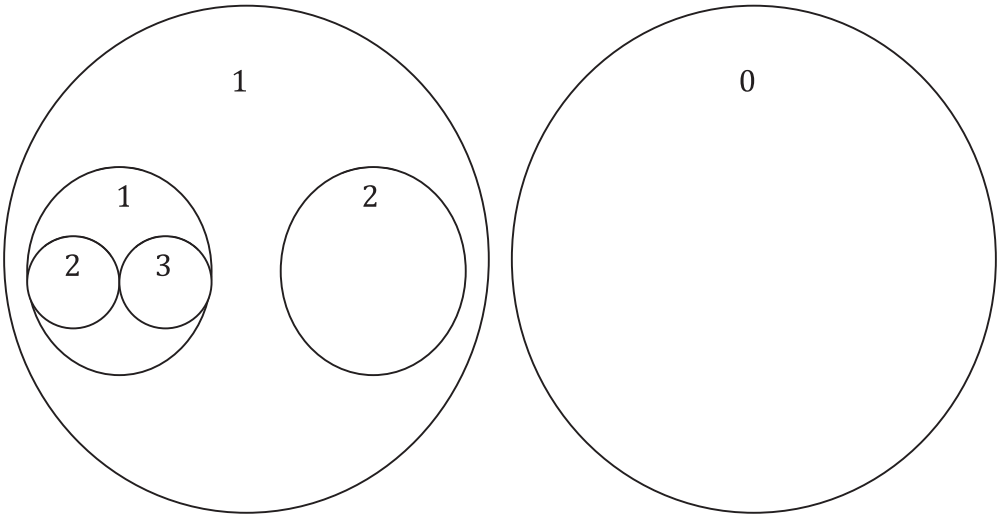
Although the use of place described thus far seems to call for a relatively simple form of modeling, for example, a concentric model (Figure 2). Such astructural model requires a concept of gradation or relativity – all places, including the world are part of a unified, non-oppositional system with a relative difference between each of the places. Relative

models create the possibility if not the probability of transformation. Within such a model the distinction between Israel and the nations becomes equally fluid and potentially transformable.

Concentric models, however, do not take into account key symbolic markers. First, they do not properly account for the intrinsic definitions of humanity (and animals) employed within the text. The definitions of humanity employed in the text are not relative and potentially transformative, they are essentialist and intrinsic, based on clear genealogical definitions. The Israelites are thus genealogically defined as Israel – on the one hand they are therefore not the nations, and the other (as a whole) not the Levites (and certainly not the Cohanim). The distinction between these categories is absolute and unbridgeable – within the Israelite model the Israelites can never become Levites and equally never become Cohanim. On this basis movement and transformation suggested by the concentric model are structurally illogical.

Purity poses similar problems for the concentric model. This issue is particularly seen at the boundary between the Camp and the world. The Camp and the world are clearly distinguished as pure and impure – with purity as being the defining characteristic of the Camp and sacred place and an intrinsic distinction between the Camp and the world. This distinction is highlighted by the Israelite rules relating to dealing with the pure and the impure. Anything considered impure must be excluded from the camp, as for example with the case of the Leper or individuals in other states of impurity (Lev. 13). Outside the Camp, the category of pure becomes meaningless. The world outside the Camp is the realm of impurity and the absence of the sacred, not in a relative sense but in an absolute sense.

Thus, the concentric model, while in part representing the shape of the camp, and the winnowing of humanity to a sacred centre, through its implication of relativity and transformation, does not account for either the intrinsic distinctions nor the oppositional relation between the pure and the impure. This suggests that the model of place must include within it an oppositional aspect, and a clear rejection of relativity. Such an oppositional model has been suggested for Israelite food rules (Kunin: 2004: 29-103).



0 The World/Nations
1 Camp/Israel
2 Court/Levites
3 Holy/Cohanim
4 Holy of Holies/High Priest (not included in diagram)

Figure 3: Oppositional Model of Centralized Sacred Place

The distinctions among/between the categories of people with differential access to sacred place can best be understood with this type of oppositional model (Figure 3). Each of the categories, be it space or people is intrinsic and unbridgeable and set in structural opposition to categories at the same level of cultural salience.

This common structural logic is consistent with structuralist theory, which suggests that a system would be defined by a consistent structural models and that recapitulating levels would all derive from the same unconscious structural equation. The structural logic of recapitulating systems is expressed in the following equation

$A_{(x)}: B_{(y)} \cong B_{(x)}:Y_{(A-)}$

This equation allows us to understand the different levels of a recapitulating structural system. The first half of the equation represents the operative layer (that is the known, upper layer of the system). In this case the opposition between Israel and the Nations. B represents Israel and A the nations. They are respectively defined by the functions (y) pure (presence of god) and (x) impure (absence of god).

The recapitulating underlying structure found in the centralized model, in both its variants, brings together two key aspects of Israelite society, sacred geography and social structure (and the nature of being an Israeite), which through their identity in structure mutually validate each other.

The centralized model, however, is not the only model of Israelite sacred place. The biblical text, perhaps due to its complexity – reflecting the internal complexity of Israelite society both synchronically and diachronically includes several different and perhaps competing models. The varient models are based on differing ideologies and theologies and often present very conflicting and mutually distinctive understanding s of sacred geography and the place of humanity and Israel within that sacred geography.

M. Haran, for example, has argued for a second **simple dynamic** model (my terminology) (Haran, 1960: 50-65). He suggests that evidence for this form of sacred place is found in Exodus 37.7, Number 12.4 and Deuteronomy 31.14. Rather than focusing on the Tabernacle and the Camp, this model focuses solely on the Tent of Meeting. While the Tent of Meeting is often associated with the Tabernacle, Haran's argument suggests that it was separable, and itself the basis of a different understanding of the relationship between god and the world. He suggests that the place of the Tent was outside of the Camp in the wilderness. His argument suggests differences in the model on two levels: conscious and unconscious.

The conscious level is essentially theological and ideological – presenting a different model of piety. He argues that the Tent represents a prophetic understanding of piety, which is focused on the individual relationship with god, while the centralized form is based on a more archaic or hierarchical communal understanding of piety. The unconscious level is structurally related to this, moving the relationship of god away from Israel, as represented by the move from the centre of the camp, and creating a relationship between god and humanity rather than merely Israel. Within Haran's model the oppositionality characteristic of the centralized model is retained. The Tent is placed in the wilderness, which is now defined as no man's land, and thus is set in opposition to all forms of society or nation.

Haran's model raises important questions about the nature of underlying structure and particularly its relationship with meaning and ideology; his model suggests a way of understanding the world that appears to be diametrically opposed to that found in the centralized model. This raises the possibility that his model is based on a differing structural equation.

The centralized model's placement of the Tabernacle in the centre of the Camp is based on the recapitulating model of structure, in which each level of the hierarchy is both intrinsic and set in structural opposition to the other category at the same level of categorization. This model at one end can be seen as having an opposition between god and man (in the person of the High Priest) and at the other one between Israel and the Nations, each based on intrinsic non-bridgability.

Haran's alternative simple dynamic model has only two categories – the divine represented by the Tent and humanity/Israel represented by the Camp. Due to the placement of the Tent in no man's land there is no structural or logical basis for making a distinction between Israel and the nations. This association fits closely with Haran's general argument that this model is more representative of a prophetic understanding than a priestly understanding.

In comparing the two models there are clear distinctions in relation to theological, ideological and sociological considerations. On the level of underlying structure these differences are based on emphasis rather than structural transformation. The centralized model emphasizes a strong view of opposition, with no mediation or transformation between categories. The categories are strictly bounded and unbridgeable. The recapitulation, which allows the development of hierarchy and

complexity, is merely the extension of the same structural equation to different levels of categorization. The structure at all levels can be characterized by the simple oppositional equation A – B. The simple dynamic model is also based on the oppositional relationship between two categories. Within this model there is no necessity or room to make distinctions within humanity, as the level of categorization is on a higher more abstract or generalizable level. Like the centralized model the simple dynamic form is based on the structural equation A – B. Thus, both variants arise from the same cultural matrix, and can be seen as competing uses of the structure, rather than alternative structures.

Alongside the centralized models discussed to this point, the bible also includes decentralized multiple forms. Singularity is perhaps the key distinctive feature of all variants on the centralized model – whether represented by the Tent, the Tabernacle or indeed the Temple in Jerusalem. The authoritative presentations of this model present this singularity without challenge – sacred place is defined by the singular Tent/Tabernacle/Temple with the Ark of the Covenant at its heart – the Ark is singular and sacred place is singular.

The decentralized multiple variants are based on multiplicity rather than the singularity highlighted to this point. Although at some point the centralized model is set in opposition to the multiple variants, other biblical texts including Judges and Kings attest to both the existence of or the establishment of other holy places existing simultaneously.

While the centralized model emphasizes in different ways the relationship between to large abstract categories, Israel/nations, divine/human in the multiple model these larger more abstract oppositions seem of less concern than internal dynamics. The text of 1 Kings 12, for example, suggests that rather than validating a singular Israelite identity, the sacred places are utilized to validate more localized identities, often or always in opposition to other localized (Israelite) identities.

The process is diagramable using a segmentary opposition genealogical model –with local sacred places creating specific local identities, and more significant places bringing together increasingly larger communal entities, in opposition to similar communal entities. This explanation of the system is important as it suggests that the underlying structure of the multiple model is identical to that of the centralized model – the difference being that the centralized model emphasizes the highest level of opposition and the multiple model the lowest. From the outside a model utilizing segmentary opposition would appear to be unified; from the inside it would appear to be multiple and decentralized.

To this point our discussion has focused on internal complexity. I have developed the discussion in some detain, both to highlight the nature of the analytical methodology and to emphasize the subtly of the differences at the structural level, and the potential significance of the differences at the narrative level. While all three examples dicussed are based on the same underlying structure, they have significantly different cultural ramifications. In each case they reflect different forms of emphasis and privileging which, although not leading these cases to the transformation of structure, they suggest the possibility of multiple small changes in emphasis and thus provide possible avenues for either evolutionary or catastrophic transformation.

Section Two: External Boundaries, solid and fuzzy.

This section focuses on a second key potential mechanism for transformation, the impact and nature of boundaries between cultural communities. Our discussion works on two interrelated levels: an abstract, huristic discussion of different structural models and the nature of the boundaries that they (conceptually) create. An a brief discussion of an ethnographic case study in which the boundaries between communities are the most salient feature of cultural identity and in which variants on the structuring of these boundaries leads to differential emphasis and potential structural transformation.

Structuralist analysis suggests that there are three ideal forms of structure, with all actual structures varying the three forms in different ways: negative, neutral and positive. Each one of these structures is primarily concerned with the categorization of content within a particular cultural context, but may also play a role in the permeability or perception of permeability of boundaries between or among different cultures.

The first form of relation is a negative relation (indicated as '-'). This relation indicates that the two categories are distinct and unbridgeable. If information x is in category A then it will never be in category B and vice versa. This is sometimes described as an oppositional relation. The second form is neutral (indicated by 'n'). In this form there is an overlap between the categories, indicating that some information in category A will also be in category B. The nature of the overlap and the value placed on movement between the categories will vary depending on the particular culture or ethnographic context. The final form is positive (indicated by '+'). In this form there are separate categories, for example A and B, but the content of the categories is identical, that is, anything in category A will also be in category B.

The biblical material discussed in the first part of this paper is based on a negative relation – all significant cultural content is in one of two opposing categories, with no content moving between categories. The neutral and positive categories are closely related. In effect the positive category is the logical extension of the neutral category. The dynamic aspect of these categories relates to the degree of mediation between the two primary categories – in this context mediation is both the degree of overlap between the categories, and the cultural understanding or valuation of that overlap.

Most discussions of categories and their relations have focus on their implications for the creation and communication of information within a cultural system. Discussions recent anthropological discussions no longer see cultures as defined and bounded entities with relatively strong boundaries between neighboring cultures. Ethnographic evidence suggests that this concept of boundedness is largely artificial. The boundaries between groups are fuzzy—overlapping and moving into each other. Thus, the coming together of different structural models, particularly but not exclusively at these fuzzy boundaries, may be a strong factor influencing the possibility of structural change. In the context of our discussion here the nature of these boundaries are significant on two interrelated levels, the narrative and structural. The narrative level is significantly more permeable than the structural, with content moving and being transformed between neighboring communities. Although this content, however, is restructured as part of the move between cultures, the boundary state requires a degree of indeterminacy and negotiation (conscious and unconscious) and thus may impact on the degree of permeability in relation to structure.

The different forms of structuring also impact both the reception of content (that is, the conscious attitude toward such reception as well as the degree and speed of reception). These are directly related to both the degree of mediation and the cultural attitude toward mediation. On the structural level, the boundaries themselves, while always being somewhat fuzzy, become increasingly fuzzy again in direct relation to the degree and attitude toward mediation. Thus, the negative relation leads to the least permeability while a positive relation may lead to the highest degree of permeability. Although in all of the cases the material is restructured,

those systems with the most permeable boundaries may also be those in which structures are more open to significant transformation (that is more than the nuancing or emphasis seen in the biblical material explored above).

The impact of different forms of structure, as well as the variability found at cultural boundaries is illustrated by material from crypto-Jewish ethnography. Crypto-Jews are individuals who believe that their ancestors were forced to convert from Judaism to Catholicism in Spain and Portugal in the 14th and 15th centuries. Despite this conversion, some aspects of Jewish identity and practices may have been passed down and have become the inheritance of their descendants living today. Modern crypto-Jews are the perfect example of fuzzy cultural boundaries. They sit on the boundary between a number of different communities, for example, Spanish, Mexican, American, Jewish and Catholic, with this final boundary catholic/Jewish being the most salient. Individuals and groups within the crypto-Jewish community and are constantly negotiating these boundaries in the ongoing process of the construction of self- identification. Their rituals and practice created through the process of bricolage enabled by the differentially fuzzy boundaries are a fertile ground for cultural improvisation.

Within the crypro-Jewish community, due to its complexity and particularly to the fact that it is inherently a boundary phenomena, a we find a complex set of variant structures, with each of these structures both serving an internal role, in defining and categorizing content, but perhaps an even more important role in defining the nature of the boundary between the different cultural/religious communities.

Three main ideal types of structures are present. On one end, there is an oppositional structure, in which Jewish cultural elements are privileged and Catholic/Christian elements are deprivileged. The structure is essentially oppositional, but has some mediation, indicated at the narrative level by the presence of Catholic/Christian content. The mediation, however, has a negative valence. This structure is associated with a relatively impermeable boundary (at least in self conception). While some content moves across the boundary, its negative ideological status reduces its cultural status and valuation – this movement both reflects and validates the mediation inherent in the structure. The fact of structural mediation itself, reflects a change in normative Jewish structure (in which mediation is highly problematic) and may be a result of the boundary state, and indeed be a feature inherent to boundary states in general, that is, that they will by definition include an aspect of mediation, however, it is viewed and valanced.

The second form of crypto-Jewish structure again has two oppositional categories, with the Jewish category being privileged over and against the Catholic/Christian category. In this form, however, there is much more significant mediation between the categories – with the mediation being both conscious and unconscious. This structure is associated with a permeable boundary (both internally and externally) and significant movement of content across the boundary. While the content that moves is not privileged it is also not strongly negatively valanced. The structure itself is transformed relative to the 'normative' Jewish structure, and shares significant features of 'normative' Christian structure.

The third form is in effect the inversion of the second. Again, two categories are present, in this case, however, it is the Christian category that is privileged rather than the Jewish. The significant mediation between categories remains. The difference between the two is that the mediation is positively valanced, thus the material that moves across the boundary is considered positive, if not valued or emphasized to the same degree as the Catholic/Christian material.

The crypto-Jewish ethnography suggests that cultural boundaries are both highly complex and fluid – with differing degrees of permeability. The fundamental characteristic of all of the

structures, however, is that despite the variability, which is much more significant that the variability discussed in the first part of this paper, the underlying structure remains largely the same – in each case the there are oppositional categories, with mediation. The difference in this case is, thus, one of privileging and the valance of the mediation. The category privileged is not a structural difference, as structure is abstract and contentless; the structure remains the same whether the Christian or Jewish material is privileged. Similarly the valancing and degree of mediation are not significant structural differences – being relative rather than intrinsic.

In the first two sections of this paper we have explored tow areas of variability in structure, and thus two possible locations for mechanisms of structural transformation. While in each case variation is found, in both the degree of structural integrity is retained. This retention of integrity has potential implications for the question of the nature of transformation as evolutionary or catastrophic. The absence of structural change via either internal differentiation or boundary variation suggests that structural transformation is not essentially gradual. It clearly relies on the accumulation of gradual changes and deformations, but the actual structural transformation from one structure to a new structural form appears to be more dramatic and catastrophic. This final aspect of our argument is developed further in a brief discussion of the change from Israelite to Christian structure.

Section Three: Opposition to Mediation a catastrophic development

Structuralist analysis of Israelite material, from throughout the Hebrew Bible, as suggested above, demonstrates the presence of a persistent underlying structural equation A – B, that is, two oppositional categories characterized a negative relation and a rejection of mediation. While, as indicated in the first section of our argument this structure has a degree of variation, this is found primarily on the narrative or ideological use of the structure, and less significantly on differential forms of emphasis. This structure can also be demonstrated to be substantially or entirely maintained in the rabbinic texts, all written or edited in the period subsequent to the birth of Christianity.

Structuralist analysis of the Gospels indicates a substantial structural transformation, that moves beyond a change in emphasis or privileging, to a new structural equation. The texts are based on an A n B equation, that is, one in which there are two categories that have significant actual or potential overlap. The relation between the categories and the valancing of the mediation is positive. Within this structure, content moves from one category to the other, and movement into the privileged category (e.g. Christian) is positive.

The Christian structure, at least arguably, arises from the same Israelite cultural matrix, that is, the same matrix and cultural context as the Rabbinic structure, and thus represents a substantial structural transformation. The difference in the two structures is best illustrated in their understanding of religious/cultural identity. The Israelite/Rabbinic oppositional structure with negative mediation gives rise to an intrinsic notion of self – based on genealogical descent. In principle one must be born an Israelite to be an Israelite and there can be no movement from non-Israel to Israel. The Christian mediated structure is based on a non-intrinsic definition of self. Christian identity is based on faith rather than birth – faith becomes the mechanism through which identity is established and by which individuals can move from the non-privileged to the privileged categories.

The significant question for our discussion is whether this transformation in structure was gradual or catastrophic. A second question relates to the issue of catalyst – this is significant particularly if the structural transformation was catastrophic. The evidence discussed in the first two sections suggested that structural integrity is high, whether it is found in the strongly oppositional Israelite material, or even the more fluid and mediated crypto-Jewish material. In both cases despite variation structural integrity was maintained, and intra-cultural

communication was retained. Little evidence is found for structural transformation that moves beyond emphasis or privileging. This suggests that although structure is fluid and variable, these variations do not undermine the integrity of the structure. Therefore significant structural transformations of the kind found in relation to the move from Israelite to Christian, are likely to be catastrophic rather than gradual. The structural change highlights a change in meaning and communication, which once it is established would make it difficult for the communities based on them to remain together – it is therefore perhaps not surprising that the break between the two traditions occurred in a relatively short space of time and was (from the perspective of both communities) complete. This break is very different from that between the various groups that composed inter-testamental and early rabbinic Judaism in which the groups although based on substantially different ideologies and practices continued to intermarry and function as a single cultural community.

The cultural and structural transformations leading to the birth of Christianity may arise from both changes in emphasis and privileging in Israelite structure and the addition of significant and powerful influences with the Hellenization of Palestine first under Greek then Roman rule. This external force, with its impact on cultural boundaries, may have provided the catalyst leading to the catastrophic transformation discussed here.

In the context of crypto-Judaism, discussed in the second section, we find a similar catastrophic transformation and an external catalyst. Crypto-Judaism emerged in the 14th and 15th centuries from the Sephardic or Spanish Jewish community. Analysis of the Sephardic texts and practices indicates that despite significant influences from both Islamic and Christian cultures, the underlying structure remained strongly oppositional with negative mediation, that is, the structure found in biblical and rabbinic material. The crypto-Jewish community emerged in a very short space of time (around 100 years) due to changes in policy towards the Jewish community and significant patterns of forced conversions. Although it is difficult to determine the exact time that structure changed, crypto-Jewish material available is based on the mediated structures discussed in the second part. This mediated structure is similar to Christian structure, but in most cases privileges the Jewish content and has a greater degree of oppositionality. Although this transformation may have been facilitated or enabled by changes in emphasis developed during the long period of positive interaction with Islamic and Christian communities, the significant structural change is directly attributable to the external catalyst provided by the forced conversions.

In this paper we have briefly touched on three aspects of structural variability and transformation. The first two parts focused on internal and boundary conditions of variability. They indicate that while variability is found in both, with boundary variability being particularly strong, in both cases essential structural integrity is maintained. The third section, focusing on structural transformation suggests that at least in the two cases discussed significant structural transformation appears to be catastrophic rather than gradual and require some form of external catalyst. This paper, however, leaves a significant question for future research, that is, is there also a form of significant structural transformation that is gradual and evolutionary. While it is clear that all cultures transform over time, it is not clear that this is associated with significant structural transformation.

Nov.29 (oral presentation)

15

Mining Metabolic Pathways through Gene Expression

Hiroshi Mamitsuka
Bioinformatics Center, Institute for Chemical Research, Kyoto University

Summary: An observed metabolic response is the result of the coordinated activation and interaction between multiple genetic pathways. However the complex structure of metabolism has meant that a complete understanding of which pathways are required to produce an observed metabolic response is not fully understood. We propose an approach which can identify the genetic pathways that dictate the response of metabolic network to specific experimental conditions. Our approach is a combination of probabilistic models for pathway ranking, clustering and classification. First we use a non-parametric pathway extraction method to identify the most highly correlated paths through the metabolic network. We then extract the defining structure within these top ranked pathways using both Markov clustering and classification algorithms. Furthermore, we define detailed node and edge annotations which enable us to track each pathway, not only with respect to its genetic dependencies, but also allow for an analysis of the interacting reactions, compounds and metabolic sub-networks. We show our approach identifies biologically meaningful pathways within two microarray expression datasets using entire metabolic networks.

1. Introduction: Essential and fundamental cellular processes are represented by metabolic networks, and their dynamics are the product of complex interactions between genes, proteins and enzymes and metabolites. Since the introduction of high-throughput technologies for measuring expression of around 10,000 genes simultaneously, the expression signatures of metabolic networks have been extensively analyzed. This mixture of network size and complexity is sufficient to hide the key pathways which define the response of the metabolic network to external stimuli. Consequently, models of global metabolic networks are required to identify the specific pathways that are driving an observed metabolic response.

A considerable amount of research has been undertaken to identify the genetic factors which dictate the function of metabolic networks [1, 2]. Of particular note are methods relating to gene set enrichment analysis (GSEA) [3] which seek to identify known groups of genes that have a non-random response to an external stimulus. GSEA methods identify important groups of genes by first ranking all genes using relatively simple test statistics such as *t*-tests or correlation coefficients, and then testing to see if specific groups are at the top or bottom of the ranked list. However, GSEA methods rely on the structure of simple test statistics which do not explicitly use the known networked structure of genes. Furthermore, they can only indicate which large prespecified groups of genes are important and do not allow for partial responses where only some genes within these known groups are related the observed response.

Probabilistic network models, having been developed in machine learning, such as those described in [4, 5] can overcome the limitations of GSEA and use feature selection to identify functional network components. At the same time these models enable to enforce these features to be logically connected within the metabolic network. However these methods require an assumption of a discrete gene expression distribution that may not completely reflect the underlying biology.

We propose an approach which conceptually lies between GSEA and probabilistic models as we assume very little about the structure of the gene expression data but enforce the identified components to be logically connected within the network. We propose a combination of three

complementary methods, having proven to be successful in analyzing small metabolic sub-networks. Firstly, we use a non-parametric pathway ranking method [6], and perform an exhaustive search to identify the top *K* most coordinated genetic pathways in response to specific experimental conditions. Our path ranking method assumes that the functional components of a metabolic network will possess a highly correlated pathway structure. Then if any functional components exist the top ranked pathways will be a clustered list of small pathway variations through these components. Pathway ranking is similar to GSEA however it explicitly uses the network structure, does not require the specification of prior groups of genes, and makes no assumption on the distribution of the gene expression.

Pathway ranking has been shown to extract biologically meaningful pathways in small metabolic sub-networks [6]. However as the network size increases, to ensure we are extracting all biologically relevant structure, we must also increase the number of pathways to be extracted. However, extracting large numbers of pathways, in the order of 1000's, prevents an easy interpretation of the result. Therefore extending pathway ranking to global metabolism requires further tools to identify the defining structures within the resulting pathway list.

To identify the defining features within the set of top ranked pathways we propose both a clustering and a classification algorithm. Both proposed algorithms exploit the natural Markov structure of a pathway. The pathway clustering algorithm is 3M [7] which identifies pathways which possess the same underlying sequence of functional genes. The pathway classifier, HME3M [8] employs the same framework as 3M but restricts the pathway search to identify those pathways which define a specific experimental condition. Both 3M and HME3M have been previously shown to identify biologically significant pathways on metabolic sub-networks [7, 8]. However these previous implementations employed a gene activity definition which required the extraction of all possible pathways between single start and end compounds within the network. Clearly this approach of extracting all possible pathways through global metabolic networks is infeasible. Therefore by using the pathways identified by the path ranking method we upscale both 3M and HME3M to analysis of global metabolic networks.

Both 3M and HME3M provide a probability estimate that each edge is a member of an identified functional component within the metabolic network. Furthermore we define detailed node and edge annotations which allow for an analysis of each pathway with respect to its core genetic dependencies and can scale up to include an analysis of interacting reactions, compounds and metabolic sub-networks. These annotations allow for the analysis of the structure within each identified functional component to be performed at multiple biologically meaningful resolutions. We show the combination of pathway ranking with Markov clustering and classification models is scalable to the analysis of complete metabolic networks. The results will highlight the flexibility of our combined approach and show that it identifies biologically meaningful pathways within real microarray expression data in both unsupervised and supervised settings.

2. Methods:

2.1. Defining Metabolic Pathways

In metabolic networks the same gene can be found in multiple locations within the network, can catalyze different reactions and therefore can possess multiple biological functions. When defining a pathway through a metabolic network the location of each gene denotes a specific function and must be precise. We define the specific location of each gene within the metabolic network through node

and edge annotations extracted from the KEGG database [10]. We define metabolic networks with nodes corresponding to genes and edges connecting two nodes. These nodes and edges are annotated as in (1) below. Each gene is annotated by its gene code (G), reaction (R) and KEGG pathway membership (P). Additionally, as the edge between two genes connects two reactions within the metabolic network, it is identified by the first substrate compound (C_F); the product compound from the first reaction (C_M); final product compound (C_T); and the final KEGG pathway membership of C_T (P_T). These annotations allow for transparent tracking of each pathway through the entire metabolic network:

$$\text{nodes} := (G,\ R,\ P); \qquad \text{edges} := (C_F, C_M, C_T,\ P_T) \tag{1}$$

We can further define a genetic pathway through a metabolic network to be a connected sequence of genes g that extend between specified start (**s**) and end compounds (**t**) as in the following (2).

$$\mathbf{s} \ldots \rightarrow g_k \rightarrow g_{k+1} \rightarrow \ldots \mathbf{t} \tag{2}$$

In (2), each edge is annotated by label shown in (1) and attached with $f(g_k, g_{k+1})$, which is a function indicating the strength of the relationship between g_k and g_{k+1}.

Our pathway definition requires the specification of start **s** and end nodes **t** which are entry and exit compounds of the metabolic network. In previous analysis we defined these nodes to be specific compounds of interest within a small metabolic sub-network [7, 8]. However these do not need to be single compounds. Therefore, to extend our original approach to analyze global metabolic networks we define **s** to be all compounds with edges only leading into the network and similarly **t** to be all compounds without edges leading into the network. Additionally we note a slight limitation of our pathway approach for the case where multiple substrates are combined by a single reaction to produce multiple products. In this situation all possible pairwise combinations of the multiple substrate and products are added to the network and then treated separately by the path ranking procedure.

Furthermore each edge is weighted by functions, $f(g_k, g_{k+1})$, which measure the strength of the relationship between g_k and g_{k+1}. The specification of the weight function is flexible with the condition that increasing values of the f indicate stronger relationships between g_k and g_{k+1}. As the accuracy of f is paramount to the path ranking method in this work we define f to be median Pearson's correlation coefficient between g_k and g_{k+1} over 100 bootstrapped replications.

A key assumption of our pathway definition is that correlated gene expression directly relates to the function of a metabolic network. It should be noted here that it is not the mRNA as measured by microarray analysis that performs the metabolic functions but the related proteins. Additionally, it is well established that mRNA expression do not always correlate well with protein abundance as it ignores key biological mechanisms and features such as post-translational modification and sub-cellular location [11]. However, in general, highly abundant proteins are also likely to show high mRNA expression levels [12]. Although the correlation between global gene expression and protein abundance is unclear, by identifying paths of maximum correlated gene expression we are focusing only on regions within the metabolic network where the expression signal is strong and are therefore more likely to possess a specific biological function.

2.2. Pathway Ranking

Probabilistic pathway ranking identifies the most probable paths through a metabolic network between

specified **s** and **t** by solving a *K*-shortest and loop-less path problem on a weighted network [6]. Our path ranking method is a non-parametric method as it does not specify the functional form of the edge weights, $f(g_k, g_{k+1})$, but instead considers the Empirical Cumulative Distribution Function (ECDF) over all edge weights within the network. In this work as we define the edge weights to be the correlation between two genes and the ECDF is simply a probabilistic rank for the most positively correlated genes. Therefore path ranking is identifying the top *K* pathways of maximal correlation through the metabolic network. Furthermore the probabilistic nature of the ECDF edge weights allow for a significance test to determine if a path contains any functional structure or is simply a random walk [6].

Due to the high levels of redundancy within metabolic networks it is likely that any pathway ranking procedure will be biased towards short paths consisting of the same gene (Note that this problem of the bias to short paths can be partially solved by computing p-values [13]). Therefore, to ensure that we extract informative paths through global metabolic networks we include two control parameters. Firstly, a parameter to control the minimum number of genes can be set to remove the numerous small and biologically uninteresting pathways from the pathway set. No maximum path length is set. Secondly, biological redundancy creates chains of reactions which are catalyzed by similar or identical gene sets. Therefore, there are multiple edges within the network where the same gene is connected to itself. As the correlation along these edges will be 1 it is clear that pathways consisting largely of the same gene will dominate the list of extracted pathways. To reduce the effect of the same gene edges we define a user specified penalty ρ on all edges which connect the same gene. We assign the edge correlation, $f(g_k, g_{k+1})$, for all same gene edges to the specified ρ value. Setting ρ allows for explicit control over the diversity of genes selected within the extracted pathway list.

We examined the scalability of the path ranking procedure to global metabolism, by using the currently available data of metabolic network and microarray [14] and a gene penalty of ρ=0. Experimental results showed that the maximum running time of the path ranking method was less than 850s (< 15 minutes) for a network with 2,828 nodes and the minimum path length of 10. This experiment was performed on an desktop PC (Intel Core i7 2.66 GHz CPU,12 GB of RAM).

2.3. 3M Pathway Clustering

The 3M Markov mixture model [7] provides the core framework for both our pathway clustering and classification models. 3M explicitly uses the Markov structure over all extracted pathways to identify the functional components. The 3M model identifies *M* key functional pathway components through a mixture of first order Markov chains for x = (s … → g_k → g_{k+1} → … t):

$$p(x)=\sum_{m=1}^{M}\pi_m p(s|\theta_{1m})\prod_{k=2}^{K}p(g_k,\text{label}_k|g_{k-1};\theta_{km}) \tag{3}$$

where π_m is the probability of each component. The probability of each component is computed by multiplying probabilities corresponding to steps over each path traversing the edge label_k linking genes g_k and g_{k+1}.

Probability parameters of the 3M model can be efficiently and conveniently estimated by an EM algorithm [6]. The result provides components, each corresponding to a cluster of frequently observed pathways that have a similar structure. Parameters for each gene within each component directly correspond to the importance of each gene within each identified functional component.

2.4. HME3M Pathway Classification

An extension to the 3M for classification is available through the HME3M model [8, 9]. HME3M

uses a Hierarchical Mixture of Experts (HME) [15] to create a classification model directly from the 3M model. To supervise 3M an additional term is added to each functional component within (3) to include information from known experimental groups,

$$p(y|X) = \sum_{m=1}^{M} \pi_m p(y|X, \beta_m) \prod_{k=2}^{K} p(g_k, \text{label}_k | g_{k-1}, \theta_{km})$$

where y is a binary response variable and X is a binary matrix where each column is a gene, each row is a pathway, and a cell value of 1 indicates the inclusion of a particular gene along a specific path. The parameters are estimated simultaneously with an EM algorithm [9, 10]. The additional term generates a classification model which takes as input the binary pathway matrix X weighted by the EM component probabilities and returns the posterior probabilities for classification of the response variable y. To ensure a scalable and interpretable solution, HME3M uses a penalized logistic regression (PLR) for each component classifier.

The goal of HME3M is to identify a set of pathways that can be used to classify a particular response label y_l. However, as we know the response variable a priori we can optimize performance of the HME3M model by directing the path ranking algorithm through network components which are differently expressed over the response labels. This can be done by normalizing the ECDF edge weights of the path ranking method across all response labels. To do this normalization step, we first compute the ECDF edge weights for each label independently and obtain $P_{EW}(g_k, g_{k-1}, y = y_l)$. Normalizing exploits the probabilistic nature of the ECDF edge weights, P_{EW}, and divides each $P_{EW}(g_k, g_{k-1}, y = y_l)$ by the sum over all response labels. We then use P_{EW} as the edge weights and extract the K most likely pathways which are specific to each response label. The process of normalizing the edge weights will highlight edges that display a difference in correlation over response labels. Normalizing the edge weights is useful for supervised analysis where the goal is to identify pathways specific to each response label, however it is not suitable for unsupervised analyses where the goal is to identify clusters of pathways with a similar structure.

3. Experiments:

We evaluate our combined methodology on two microarray datasets constructed with differing experimental objectives. The first microarray is the benchmark Gasch microarray [14] which observes the genetic response of yeast to numerous environmental stress conditions. Our goal is to identify which stress conditions have similar pathway responses. The second is a microarray observing the genetic differences between equally obese patients who are insulin resistant versus those who are insulin sensitive [16]. Insulin resistance is a known metabolic hallmark of type II diabetes. Our goal in this analysis is to identify specific pathway differences between insulin resistant and sensitive patients. The microarray datasets were obtained from GEO [17] (Gasch = GSE18, diabetes = GSE121).

3.1. Clustering Yeast Stress Responses

The structure of the 3M model for clustering the Gasch yeast stress conditions is presented in Fig. 1. Fig. 1 contains two tables where the upper table is a heat map of the clustering confusion matrix with the known stress conditions on the left axis and the 3M cluster labels on the top axis. The numbers within each cell are the number of pathways for each stress condition that occur within each 3M cluster. The lower table is a heat map of the pathway structure for each 3M component. The left axis shows the KEGG metabolic pathways, compounds and reactions, the right axis shows the genes. The number within each cell is probability of each gene within each 3M component. To improve the clarity

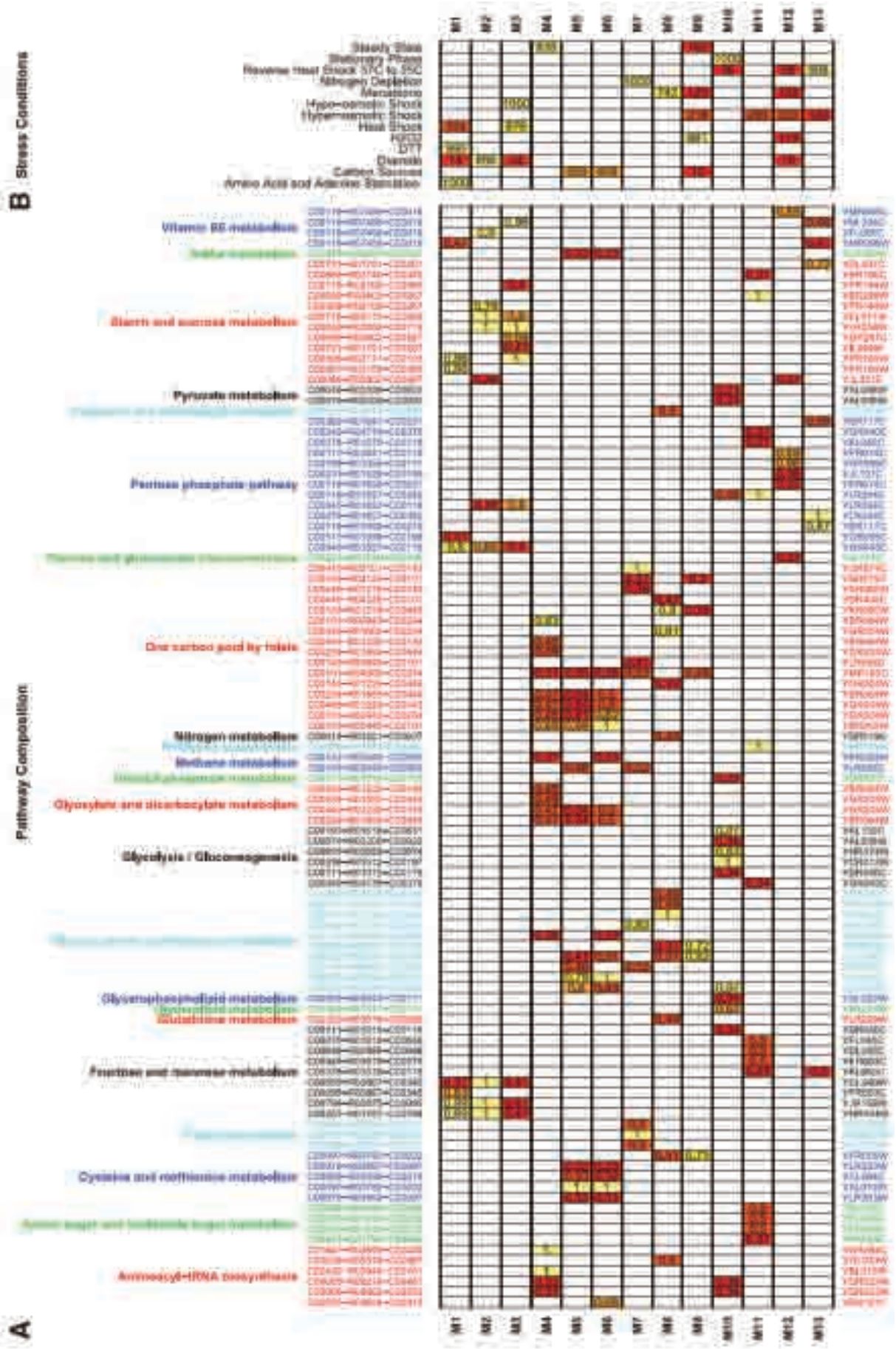


Fig. 1

of the display in Fig. 1 all parameters of less than 0.3 were removed and then only columns with more

than one valid parameter were retained. We here discuss on the first group of clusters: M1, M2 and M3 (Heat shock, Hypo-osmotic shock, DTT, Amino-acid starvation, Diamide).

The stress conditions of heat shock, hypo-osmotic shock, DTT, amino-acid starvation and diamide are grouped together as they share common components across Starch and sucrose metabolism, Fructose and mannose metabolism and the Pentose phosphate pathway. The main common reaction path is from C00267 (alpha-D-Glucose) to C00018 (pyridoxal phosphate) through C00794 (Sorbitol), C00095 (D-Fructose), C05345 (beta-D-Fructose-6P) and C00118 (Glyceraldehyde-3-P). Despite this large common element 3M is able to distinguish three sub-clusters which group DTT with amino acid starvation, hypo-osmotic stress with heat shock and lastly diamide exposure. These pathway differences occur mainly within Starch and sucrose metabolism.

The (DTT, amino acid starvation) (M1) cluster is defined by an edge from alpha-D-Glucose to C00103 (D-Glucose 1P) by the YPR184W gene. YPR184W is induced by Gcn4p, which is a key regulator expressed during amino acid starvation in yeast [18]. Gcn4p is also known to regulate genes relating to the unfolded protein response initiated by DTT exposure [19]. Therefore the similarity between the DTT and amino acid starvation stress responses is likely due to Gcn4p regulation.

The diamide response (M2) has considerable overlap with both M1 and M3 but includes key components that allow it to be clustered separately. The similarity between diamide, heat shock, DTT agrees with observation by [14]. The unique component of the diamide response is largely defined by the conversion from C00369 (starch) to C00267 (alpha-D-Glucose) by YPR184W. YPR184W can be regulated by the Yap1p gene [20] which is known to be activated during diamide exposure [21].

The (heat shock, hypo-osmotic stress) cluster (M3) is defined by the synthesis of alpha-D-Glucose from C00721 (Dextrin) or C00089 (sucrose) and not from starch as in M2. In M3, starch is converted to C00103 (alpha-D-Glucose 1-phosphate) by gene YPR160W. The stress response to heat shock and hypo-osmotic shock are known to be similar [22] and both can be regulated by the heat shock transcription factors [23]. Furthermore, YPR160W is known to be regulated by the Hog1p-MAP pathway [24] which is known activated under osmotic stress conditions and heat shock [25].

One striking feature common to all M1, M2 and M3 is the use of the gene YPR184W (GDB1). YPR184W is a glycogen debranching enzyme (reaction R02109) which catalyzes the conversion from glycogen (C00718) to starch (C00369) and then from starch to alpha-D-glucose (C00267). Although all components (M1, M2 and M3) include YPR184W within their components the function of YPR184W is different in each component. M1 uses YPR184W to convert alpha-D-glucose into starch and then interacts YPR184W with YPR160W (GPH1) to synthesize alpha-D-glucose 1P (C00103). M2 is the reverse pathway to M1 and uses YPR184W to convert starch to alpha-D-glucose and then interacts YPR184W with YHR104W (GRE3) to produce sorbitol. In contrast M3 begins with glycogen and uses YBR184W to create starch and then proceeds in a similar path to M1. Therefore the component differences surrounding the use of YPR184W identifies how starch is metabolized within the network in different stress conditions. The ability to identify multiple functions of single genes and compounds as illustrated by YPR184W is a powerful feature of using pathways to model metabolic networks.

3.2. Classification on Insulin Resistance

The key component for insulin resistance identified by HME3M is presented in Fig. 2, which shows the same component in terms of its connected pathways (A) and compounds (B). The line thickness represents highly probable pathways. It is clear from Fig. 2A that purine metabolism is the primary driver of insulin resistance. The highly probable edges within compound network (Fig. 2B) relate to

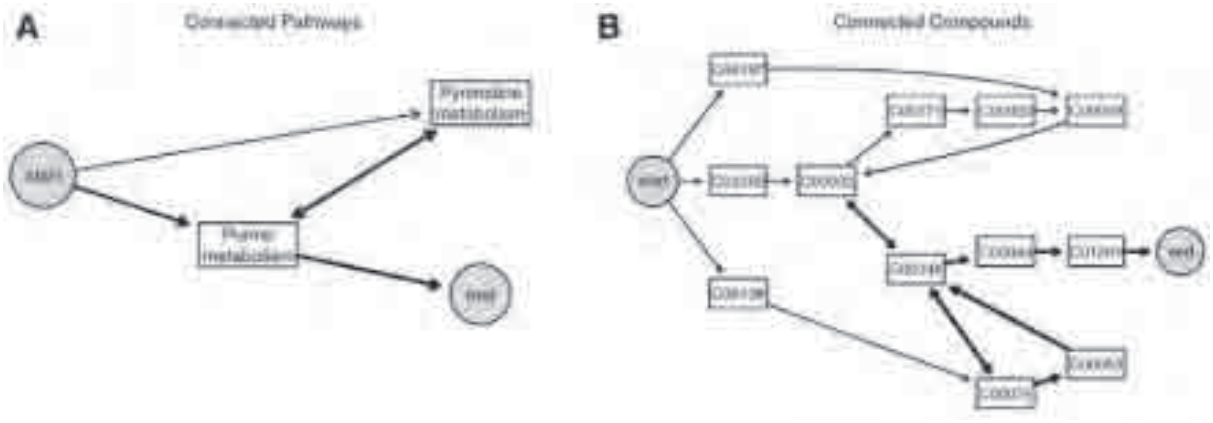


Fig. 2

conversions between C00002 (ATP) through C00046 (RNA) C00075 (UTP), C00063 (CTP), C00044 (GTP) and C01261 (GppppG). These steps are performed by a common set of reactions and genes. Although this is a common set of compounds is a major feature of the pathway structure it is not strongly related to insulin resistance but rather serves as the exit point used by all pathways through purine metabolism. The features that relate strongly to insulin resistance are compounds used to create ATP a key signaling molecule in diabetes and insulin secretion [26].

In Fig. 2B, ATP is observed to be synthesized from C01260 (AppppA), C06197 (ApppA), C06198 (UppppU) or converted to C00575 (cAMP), C00020 (AMP) and C00008 (ADP) and then back to ATP. Both AppppA and ApppA are known to have a role in insulin release [27] and have more recently been linked to diabetes [28]. Additionally cyclic AMP (cAMP) has recently been indicated in the production of GLP-1 which is a known factor in insulin secretion and type II diabetes [29]. GLP-1 is also target for the drug Byetta which is now being developed to treat type II diabetes [29]. Furthermore, the conversion from ADP back to ATP is an oxidative phosphorylation step using the nucleoside diphosphate kinase (NDK) enzyme [30]. NDK has been previously linked to type 2 diabetes [31] and abnormal oxidative phosphorylation is a known factor in type II diabetes [32].

3. Conclusions:

We have described a complete process for extracting and analyzing functional pathways within global metabolic networks and gene expression data. Our approach allows for analysis at the sub-network, compound, reaction and gene resolutions which allows for a complete picture of the metabolic response. We have shown that our combined approach extracts biologically meaningful results by identifying key genetic regulators of environmental stress in yeast and known drug targets for metabolic networks in humans. Furthermore, our approach is flexible allowing complete control over the structure and diversity of the pathways to be found and caters for both unsupervised and supervised styles of analysis. We believe these results highlight our approach to be a powerful framework for the analysis of global metabolic networks.

We note that this is a joint work with Timothy Hancock and Ichigaku Takigawa of Kyoto University. Detailed versions of this abstract appeared in [33, 34].

Acknowledgments

This work has been supported in part by BIRD of JST and KAKENHI (24300054) of MEXT, Japan.

References

[1] Karp, PD, Paley, SM, Krummenacker, M, Latendresse, M, Dale, JM, Lee, TJ, Kaipa, P, Gilham, F, Spaulding, A, Popescu, L, Altman, T, Paulsen, I, Keseler, IM, Caspi, R (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief. Bioinformatics*, 11, 1:40-79.

[2] Mlecnik, B, Scheideler, M, Hackl, H, Hartler, J, Sanchez-Cabo, F, Trajanoski, Z (2005). PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res*., 33, Web Server issue:W633-7.

[3] Subramanian, A, Tamayo, P, Mootha, VK, Mukherjee, S, Ebert, BL, Gillette, MA, Paulovich, A, Pomeroy, SL, Golub, TR, Lander, ES, Mesirov, JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, 102, 43:15545-50.

[4] Wei, Z, Li, H (2007). A Markov random field model for network-based analysis of genomic data. *Bioinformatics*, 23, 12:1537-44.

[5] Sanguinetti, G, Noirel, J, Wright, PC (2008). MMG: a probabilistic tool to identify submodules of metabolic pathways. *Bioinformatics*, 24, 8:1078-84.

[6] Takigawa, I, Mamitsuka, H (2008). Probabilistic path ranking based on adjacent pairwise coexpression for metabolic transcripts analysis. *Bioinformatics*, 24, 2:250-7.

[7] Mamitsuka, H, Okuno, Y and Yamaguchi, A (2003). Mining Biologically Active Patterns in Metabolic Pathways using Microarray Expression Profiles. *ACM SIGKDD Explorations*, 5 (2), 113-121

[8] Hancock, T, and Mamitsuka, H (2009). A Markov Classification Model for Metabolic Pathways *Lecture Notes in Bioinformatics* 5724 *(Proceedings of the Ninth Workshop on Algorithms in Bioinformatics (WABI 2009))*, 121-132.

[9] Hancock, T. and Mamitsuka, H (2010). A Markov Classification Model for Metabolic Pathways *Algorithms for Molecular Biology* 5, 10.

[10] Kanehisa, M, Goto, S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28, 1:27-30.

[11] Gygi, SP, Rochon, Y, Franza, BR, Aebersold, R (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.*, 19, 3:1720-30.

[12] Ghaemmaghami, S, Huh, WK, Bower, K, Howson, RW, Belle, A, Dephoure, N, O'Shea, EK, Weissman, JS (2003). Global analysis of protein expression in yeast. *Nature*, 425, 6959:737-41.

[13] Hancock, T, Wicker, N, Takigawa, I, Mamitsuka, H (2012). Identifying neighborhoods of coordinated gene expression and metabolite profiles. *PLoS ONE*, 7, 2:e31345.

[14] Gasch, AP, Spellman, PT, Kao, CM, Carmel-Harel, O, Eisen, MB, Storz, G, Botstein, D, Brown, PO (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11, 12:4241-57.

[15] Jordan, M and Jacobs, R (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.*, 6, 181-214.

[16] Yang, X, Pratley, RE, Tokraks, S, Bogardus, C, Permana, PA (2002). Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians. *Diabetologia,* 45, 11:1584-93.

[17] Barrett, T, Troup, DB, Wilhite, SE, Ledoux, P, Evangelista, C, Kim, IF, Tomashevsky, M, Marshall, KA, Phillippy, KH, Sherman, PM, Muertter, RN, Holko, M, Ayanbule, O, Yefanov, A, Soboleva, A (2011). NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res.*, 39, Database issue:D1005-10.

[18] Natarajan, K, Meyer, MR, Jackson, BM, Slade, D, Roberts, C, Hinnebusch, AG, Marton, MJ (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol. Cell. Biol.*, 21, 13:4347-68.

[19] Patil, CK, Li, H, Walter, P (2004). Gcn4p and novel upstream activating sequences regulate

targets of the unfolded protein response. *PLoS Biol.*, 2, 8:E246.
[20] García, R, Rodríguez-Peña, JM, Bermejo, C, Nombela, C, Arroyo, J (2009). The high osmotic response and cell wall integrity pathways cooperate to regulate transcriptional responses to zymolyase-induced cell wall stress in Saccharomyces cerevisiae. *J. Biol. Chem.*, 284, 16:10901-11.
[21] Kuge, S, Arita, M, Murayama, A, Maeta, K, Izawa, S, Inoue, Y, Nomoto, A (2001). Regulation of the yeast Yap1p nuclear export signal is mediated by redox signal-induced reversible disulfide bond formation. *Mol. Cell. Biol.*, 21, 18:6139-50.
[22] Jamieson, DJ (1998). Oxidative stress responses of the yeast Saccharomyces cerevisiae. *Yeast*, 14, 16:1511-27.
[23] Cabiscol, E, Bellí, G, Tamarit, J, Echave, P, Herrero, E, Ros, J (2002). Mitochondrial Hsp60, resistance to oxidative stress, and the labile iron pool are closely connected in Saccharomyces cerevisiae. *J. Biol. Chem.*, 277, 46:44531-8.
[24] Sunnarborg, SW, Miller, SP, Unnikrishnan, I, LaPorte, DC (2001). Expression of the yeast glycogen phosphorylase gene is regulated by stress-response elements and by the HOG MAP kinase pathway. *Yeast*, 18, 16:1505-14.
[25] Mollapour, M, Piper, PW (2006). Hog1p mitogen-activated protein kinase determines acetic acid resistance in Saccharomyces cerevisiae. *FEMS Yeast Res.*, 6, 8:1274-80.
[26] Koster, JC, Permutt, MA, Nichols, CG (2005). Diabetes and insulin secretion: the ATP-sensitive K+ channel (K ATP) connection. *Diabetes*, 54, 11:3065-72.
[27] Verspohl, EJ, Johannwille, B (1998). Diadenosine polyphosphates in insulin-secreting cells: interaction with specific receptors and degradation. *Diabetes*, 47, 11:1727-34.
[28] Rüsing, D, Verspohl, EJ (2004). Influence of diadenosine tetraphosphate (Ap4A) on lipid metabolism. *Cell Biochem. Funct.*, 22, 5:333-8.
[29] Yu, Z, Jin, T (2010). New insights into the role of cAMP in the production and function of the incretin hormone glucagon-like peptide-1 (GLP-1). *Cell. Signal.*, 22, 1:1-8.
[30] Lipskaya, TY, Voinova, VV (2005). Functional coupling between nucleoside diphosphate kinase of the outer mitochondrial compartment and oxidative phosphorylation. *Biochemistry Mosc.*, 70, 12:1354-62.
[31] Zhu, J, Tseng, YH, Kantor, JD, Rhodes, CJ, Zetter, BR, Moyers, JS, Kahn, CR (1999). Interaction of the Ras-related protein associated with diabetes rad and the putative tumor metastasis suppressor NM23 provides a novel mechanism of GTPase regulation. *Proc. Natl. Acad. Sci. U.S.A.*, 96, 26:14911-8.
[32] Højlund, K, Yi, Z, Lefort, N, Langlais, P, Bowen, B, Levin, K, Beck-Nielsen, H, Mandarino, LJ (2010). Human ATP synthase beta is phosphorylated at multiple sites and shows abnormal phosphorylation at specific sites in insulin-resistant muscle. *Diabetologia*, 53, 3:541-51.
[33] Hancock, T, Takigawa, I, Mamitsuka, H (2010). Mining metabolic pathways through gene expression. *Bioinformatics*, 26, 17:2128-35.
[34] Hancock, T, Mamitsuka, H (2010). Active pathway identification and classification with probabilistic ensembles. *Genome Inform*., 22:30-40.

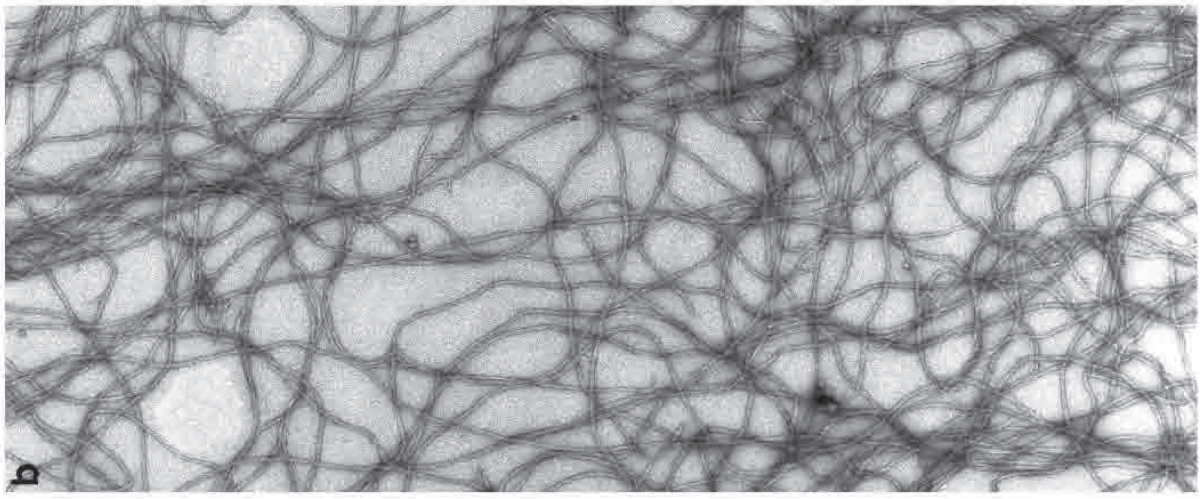
16

How shape and function of cells and tissues is determined by intermediate filaments and their associated protein chaperones

Professor Roy Quinlan
School of Biological and Biomedical Sciences; Biophysical Sciences Institute;
University of Durham

Summary: The shape and function of vertebrate cells and tissues is determined largely by the cytoskeleton. The most resilient component in the cytoskeleton is intermediate filaments, which have exceptional strain-hardening properties. These filamentous protein polymers function in part to counter the physical stresses experienced by cells and tissues. When compromised by mutation, then a plethora of human diseases can result depending on the specific intermediate filament protein involved and its tissue specific expression pattern. In this presentation I will detail the assembly of these 10nm (nano-)filaments and show how mutation compromises their function in cells and tissues using a range of *in vitro* and i*n vivo* approaches. I will show how intermediate filament function requires the association of protein chaperones. Indeed, intermediate filaments and small heat shock protein (sHSP) chaperones form a function complex that is key to counteracting stress and also propagating signals to determine whether a cell lives or dies. Intermediate filaments perform a scaffolding function in the cell stress response and indeed act as a platform for key proteins involved in the stress signaling cascade providing a potential mechanistic explanation for their key role in modulating cellular stress. It is therefore not too surprising then that protein chaperones also associate with intermediate filaments given their role in the stress response, but it was not expected that this association could compromise intermediate filament function. This we know from studying cardiomyopathies that demonstrate the tight functional interaction between desmin, the cardiac specific intermediate filament protein, and αB-crystallin (CRYAB), but which we can also demonstrate in other tissue such as the eye lens and in some specific cell types like astrocytes. Therefore in this presentation I will illustrate the inter-dependence of intermediate filaments, sHSPs and the shape and function of mammalian cells and tissues to facilitate their application in future smart technologies.

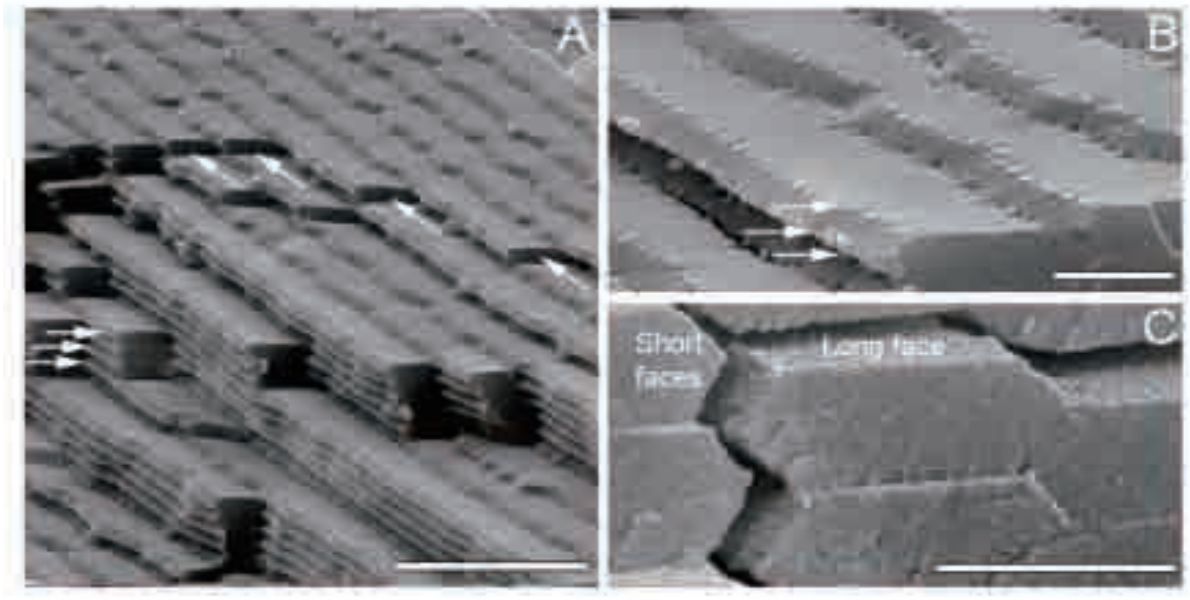
1. Introduction: The main research interests of my lab are the cytoskeleton and protein chaperones of mammalian cells, their role in the stress response and their links with human diseases. Intermediate filament proteins along with small heat shock protein chaperones are early response genes after heat shock. Each form different nanostructures – one forms 10nm filaments as shown below. The filaments



can be made recombinantly in *E. coli* allowing the assembly process to be studied *in vitro* using natural or selected mutations to disrupt or modify the assembly process. The filaments are many microns long

but exhibit this exquisite control in width. The sHSPs on the other hand form 15nm particles although the number of subunits within the protein complex of mammalian sHSPs can be variable, in archael orthologues it is tightly controlled at 24 subunits with octahedral symmetry. In both the case of intermediate filaments and sHSPs, the polymers are assembled from protein monomers. The polymers have potential biomimetic application in future smart material technologies that could include nanowires or mineralization templates to drug delivery systems, and so as a structural biologist, I am intrigued by their assembly and how this is controlled and regulated in living cells.

As a cell biologist, I am also interested in the function of the polymers and how they function in individual cells and the tissue itself. Mutations in intermediate filament proteins cause many different diseases, but it is the pheno-copying of desmin mutations by those in the small heat shock protein αB-crystallin (CRYAB) that demonstrated genetically the functional interaction between these protein chaperones and the intermediate filaments as a key element of the cytoskeleton. As an illustration of this point, the image opposite of a cell shows the immunofluorescence signal for a cell stained with antibodies to detect a sHSP. The filamentous pattern is clearly obvious – it has been colour coded for depth (blue is close to substratum) and there is little doubt there for that sHSPs are intimately associated with intermediate filaments. This can be demonstrated by immunofluoresence light microscopy and immunogold labeling electron microscopy. In my presentation I will present our evidence for this functional association between CRYAB and desmin and also several other intermediate filament proteins, GFAP and BFSP1/2. We find the involvement of the intermediate filament-sHSP complex strongly implicates an altered stress response as a common element in the aetiology of the various human diseases caused by mutations in intermediate filament proteins. Our long-standing interest in the eye lens and the precise



hexagonal profile of the lens fibre cells (see above; the cross sectional profile of lens fibre cells is hexagonal (panel C), and at lower magnification (panel A) the stacking of individual lens cells, which

can be millimeters long, resembles planks of wood as intercellular space is minimized. This is achieved by the interdigitation of neighbouring cells at their interfaces (white arrows; panel B) and by offsetting each cell to precisely organize the lens cells (panel C). The lens is therefore a perfect illustration of the functional importance of this chaperone-cytoskeleton interaction to the optical properties of the eye lens. When a key component (Bfsp2) of the lens intermediate filaments was removed from the mouse genome then this altered the optical properties of the lens and also affected cell shape, causing greater light scatter and loss of refraction. Mutations in this protein cause cataract in humans. Therefore the lens is a very good example of how cell shape and tissue function are interlinked. It also illustrates too that there are functions for this sHSP-intermediate filament complex even when stress is not overtly involved.

2. Discussion: The mechanistic and molecular details of the assembly of intermediate filaments are important areas of research. Attempts to mimic the assembly of such filaments that exhibit such precise control in width, whilst at the same time capable of such extensive assembly has yet to be achieved with synthesized peptides. For future technological application, it will be important to understand intermediate filament assembly in detail. Then the interaction between sHSPs and intermediate filament proteins is clearly an important complex with roles in many different cell biological processes. The selection of intermediate filaments over other potential chaperone clients is an equally important research avenue to pursue, how a stable complex is retained in cells. Then in terms of the functional aspects of the protein complex, we can provide evidence to the scaffolding role of the intermediate filaments within the context of a stress response, but within unstressed cells, the role of the sHSP-intermediate filament complex is not as clear cut. The data suggest that the association of sHSPs with intermediate filaments could assist in their subcellular distribution by preventing filament aggregation. The corollary of this is that perhaps this association also assists in the selection of client proteins for sHSPs. Nevertheless our data point to a very important but neglected aspect of the function of protein chaperones – namely their contribution to the assembly of protein complexes and polymers. Here the sHSPs provide a prime fascia example the fundamental importance of this function.

Reference:

1. Houck SA, Landsbury A, Clark JI, Quinlan RA. Multiple sites in αB-crystallin modulate its interactions with desmin filaments assembled in vitro. PLoS One. 2011;6(11):e25859. Epub 2011 Nov 9. PMID: 22096479
2. Chen YS, Lim SC, Chen MH, Quinlan RA, Perng MD. Alexander disease causing mutations in the C-terminal domain of GFAP are deleterious both to assembly and network formation with the potential to both activate caspase 3 and decrease cell viability. Exp Cell Res. 2011 Oct 1;317(16):2252-66. Epub 2011 Jul 2. PMID: 21756903
3. Dahm R, van Marle J, Quinlan RA, Prescott AR, Vrensen GF. Homeostasis in the vertebrate lens: mechanisms of solute exchange. Philos Trans R Soc Lond B Biol Sci. 2011 Apr 27;366(1568):1265-77. PMID: 21402585
4. Tang G, Perng MD, Wilk S, Quinlan R, Goldman JE.Oligomers of mutant glial fibrillary acidic protein (GFAP) Inhibit the proteasome system in alexander disease astrocytes, and the

small heat shock protein alphaB-crystallin reverses the inhibition. J Biol Chem. 2010 Apr 2;285(14):10527-37. Epub 2010 Jan 28. PMID: 20110364

5. Perng MD, Wen SF, Gibbon T, Middeldorp J, Sluijs J, Hol EM, Quinlan RA. Glial fibrillary acidic protein filaments can tolerate the incorporation of assembly-compromised GFAP-delta, but with consequences for filament organization and alphaB-crystallin association. Mol Biol Cell. 2008 Oct;19(10):4521-33. Epub 2008 Aug 6. PMID: 186850832:
6. Der Perng M, Su M, Wen SF, Li R, Gibbon T, Prescott AR, Brenner M, Quinlan RA. The Alexander disease-causing glial fibrillary acidic protein mutant, R416W, accumulates into Rosenthal fibers by a pathway that involves filament aggregation and the association of alpha B-crystallin and HSP27. Am J Hum Genet. 2006 Aug;79(2):197-213. Epub 2006 Jun 12. PMID: 16826512
7. Perng MD, Wen SF, van den IJssel P, Prescott AR, Quinlan RA. Desmin aggregate formation by R120G alphaB-crystallin is caused by altered filament interactions and is dependent upon network status in cells. Mol Biol Cell. 2004 May;15(5):2335-46. Epub 2004 Mar 5. PMID: 15004226
8. Sandilands A, Wang X, Hutcheson AM, James J, Prescott AR, Wegener A, Pekny M, Gong X, Quinlan RA. Bfsp2 mutation found in mouse 129 strains causes the loss of CP49' and induces vimentin-dependent changes in the lens fibre cell cytoskeleton. Exp Eye Res. 2004 Apr;78(4):875-89. PMID: 15037121
9. Perng MD, Cairns L, van den IJssel P, Prescott A, Hutcheson AM, Quinlan RA. Intermediate filament interactions can be altered by HSP27 and alphaB-crystallin. J Cell Sci. 1999 Jul;112 (Pt 13):2099-112. PMID: 10362540
10. Nicholl ID, Quinlan RA. Chaperone activity of alpha-crystallins modulates intermediate filament assembly. EMBO J. 1994 Feb 15;13(4):945-53. PMID: 7906647

17

Chemistry of low-barrier hydrogen bonds in protein active sites

Hiroshi Ishikita
Career-Path Promotion Unit for Young Life Scientists, Kyoto University
Japan Science and Technology Agency (JST), PRESTO

Summary
The photoactive chromophore of photoactive yellow protein (PYP) is *p*-coumaric acid (*p*CA). Recent neutron diffraction studies of PYP proposed that the H bond between protonated Glu46 and the chromophore (ionized *p*CA) was a low-barrier H bond (LBHB). Using the atomic coordinates of the high-resolution crystal structure, we analyzed the energetics of the short H bond by two independent methods: electrostatic pK_a calculations and a quantum mechanical/molecular mechanical (QM/MM) approach. (i) In the QM/MM optimized geometry, we reproduced the two short H-bond distances of the crystal structure: Tyr42–*p*CA (2.50 Å) and Glu46–*p*CA (2.57 Å). However, the H atoms obviously belonged to the Tyr or Glu moieties, and were not near the midpoint of the donor and acceptor atoms. (ii) The potential-energy curves of the two H bonds resembled those of standard asymmetric double-well potentials, which differ from those of LBHB. (iii) The calculated pK_a values for Glu46 and *p*CA were 8.6 and 5.4, respectively. The pK_a difference was unlikely to satisfy the prerequisite for LBHB. (iv) The LBHB in PYP was originally proposed to stabilize the ionized *p*CA because deprotonated Arg52 cannot stabilize it. However, the calculated pK_a of Arg52 and QM/MM optimized geometry suggested that Arg52 was protonated on the protein surface. The short H bond between Glu46 and ionized *p*CA in the PYP ground state could be simply explained by electrostatic stabilization without invoking LBHB.

1. Introduction

Sensing blue light is a prerequisite for organisms to be able to sustain life. Photoactive yellow protein (PYP) serves as a bacterial photoreceptor, in particular, as a sensor for negative phototaxis to blue light (1). The photoactive chromophore of PYP is *p*-coumaric acid (*p*CA), which is covalently attached to Cys69 (2). In the PYP ground state, the *p*CA chromophore is present as a phenolate anion (3-5). Absorption of a blue light photon initiates the *trans–cis* isomerization of the *p*CA region, leading to proton transfer involving the *p*CA moiety (4, 6). The PYP crystal structure revealed that *p*CA is H-bonded by protonated Tyr42 and Glu46 (Figure 1). Tyr42 is further H-bonded by Thr50. Structural analysis suggested that Glu46 is protonated and *p*CA is ionized in the PYP ground state (7, 8). Remarkably, the distance between the hydroxyl O of Tyr42 and the phenolate O of *p*CA (O_{Tyr42}–O_{pCA}) is 2.49–2.51 Å, and the distance between the carboxyl O of Glu46 and the phenolate O of *p*CA (O_{Glu46}–O_{pCA}) is 2.54–2.61 Å in most PYP crystal structures at resolution of ~1 Å (reviewed in Ref. (9)). The short distance between Glu46 and *p*CA is of particular interest because the photoinduced intramolecular proton transfer from protonated Glu46 to ionized *p*CA occurs in transition to the pR intermediate state during the photocycle (4).

Recently, using the heavy atom coordinates of the PYP X-ray diffraction crystal structure analyzed at 1.25 Å resolution, hydrogen or deuterium atom positions of PYP were assigned in neutron

diffraction analysis at 295 K (10). (note: both hydrogen (H) and deuterium (D) are called H atom in the present study. Changes in the H-bond donor–acceptor distances due to H/D substitution are subtle, 0.01 Å in NMR studies on PYP (11).). According to the neutron diffraction analysis, an H atom in the O_{Tyr42}–O_{pCA} bond was located at 0.96 Å from Tyr42 and was assumed to be an ionic H bond (10). In contrast, in the case of the Glu46–*p*CA pair, an H atom position was assigned at 1.21 Å from Glu46 and 1.37 Å from *p*CA, almost at the midpoint of the O_{Glu46}–O_{pCA} bond (2.57 Å) (Figure 1a). From this unusual H atom position, the H bond between Glu46 and *p*CA was interpreted as a low-barrier H bond (LBHB (12)) by the authors of Ref. (10). LBHB was originally proposed to possess a covalent bond-like character, thus significantly stabilizing the transition state and facilitating enzymatic reactions (12, 13). To understand the O_{Glu46}–O_{pCA} bond characteristics, the following points should be considered:

(a) H-bond length and NMR chemical shift. It was suggested that a strong H-bond results in a more downfield ^{1}H NMR chemical shift. According to the classification of H bonds by Jeffrey (14) or Frey (15), “single-well H bonds” (or “symmetrical H bonds” (16)) are very short typically with O–O distances of 2.4 to 2.5 Å and display ^{1}H NMR chemical shifts (δ_H) of 20 to 22 ppm (15). “LBHBs” (or “asymmetric H bonds” (16)) are longer, 2.5 to 2.6 Å with δ_H of 17 to 19 ppm (15). “Weak H-bonds” are further longer, with δ_H of 10 to 12 ppm (15). According to the criteria (15, 16), the O_{Glu46}–O_{pCA} bond is not an LBHB but is more likely to be a single-well H-bond in terms of the H atom position. However, simultaneously, the reported O_{Glu46}–O_{pCA} distance of 2.57 Å (10) is too long for a single-well H bond. Thus, it is not clear whether this protein has an LBHB on the basis of the H-bond geometry.

On the other hand, δ_H of 15.2 ppm was assigned to protonated Glu46 in NMR studies (11). The value of 15.2 ppm is smaller than that for single-well H bonds (20 to 22 ppm (15)) or even for LBHB (17 to 19 ppm (15)).

(b) pK_a values. According to Perrin and Nielson (17) or Schutz and Warshel (18), the definition “LBHB” is vague. Schutz and Warshel (18) concluded that LBHB cannot be defined by only the distance or strength of a H-bond, and that only energy-based evaluations can be used to determine H-bond types. In particular, the pK_a values of the donor and acceptor moieties are important in determining the energy barrier required for moving an H atom between donor and acceptor moieties (18). In original reports by Frey et al. (13) or Cleland and Kreevoy (12), it was stated that an LBHB may form when the pK_a difference between donor and acceptor moieties is nearly zero. Interestingly, it was also speculated by the authors of Ref. (10) that the pK_a values of Glu46 and *p*CA would be similar in the PYP ground state. If this were the case, the potential-energy curve of the H-bond should resemble the shape of a symmetric potential (18); this would be consistent with the H atom position at the midpoint of the O_{Glu46}–O_{pCA} bond, as reported in the neutron diffraction analysis (10) (note: Frey

considered LBHBs as “asymmetric H bonds” (16)). However, the “similar pK_a values of Glu46 and *p*CA” contradict the protonated Glu46 and deprotonated *p*CA in the PYP ground state, as suggested in a number of previous experimental studies (3-5).

(c) covalent-bond like character. In PYP, Arg52 is located on the protein surface in the chromophore region and it shields the chromophore from the solvent, which, thus far, was considered to be protonated (7) (Figure 1). In contrast, Arg52 was concluded to be deprotonated by the authors of Ref. (10) due to the absence of the corresponding nuclear density. They speculated that ionized *p*CA was energetically unstable in the hydrophobic chromophore unless a covalent-bond like LBHB was present. However, it should be noted that several polar groups exist close to the *p*CA in the PYP chromophore, e.g., Tyr42 and Thr50. In addition, the authors of Ref. (10) did not provide any explanation as to how is the existence of deprotonated Arg52 energetically possible on a protein surface where solvation energy is sufficiently available.

In the present study, we investigated how the formation of the short H bond between Glu46 and *p*CA was energetically favored in the 1.25-Å PYP crystal structure, using a large-scale quantum mechanical/molecular mechanical (QM/MM) approach (*19*). We further report a H atom position in the O_{Glu46}–O_{pCA} bond and the corresponding ^{1}H NMR chemical shifts, δ_H. We calculated δ_H for O_{Glu46}–O_{pCA} quantumchemically, with the full account of the complete PYP atomic coordinates; defining the *p*CA with the covalently bonded Cys69 and all H-bond partner residues, i.e., Tyr42, Glu46, and Thr50, as the QM region and the remaining residues as the MM region (*20*).

We also performed pK_a calculations by solving the linear Poisson–Boltzmann (LPB) equation and considered the protonation states of all titratable sites of PYP (electrostatic calculation). Note that in general, electrostatic and QM/MM calculations give consistent results, e.g., (*21*).

2. Results and Discussion

H-bond distances in PYP. The QM/MM geometry optimization resulted in an O_{Tyr42}–O_{pCA} distance of 2.50 Å and an O_{Glu46}–O_{pCA} distance of 2.57 Å (*19*), which are in agreement with the distances of 2.52 Å and 2.57 Å in neutron diffraction analysis (10), respectively (Table 1). Note that the corresponding distances are 2.50 Å and 2.59 Å in another X-ray diffraction analysis at a resolution of 1.00 Å (9), respectively. The H atom in the O_{Tyr42}–O_{pCA} bond was located at a distance of 1.01 Å from Tyr42 rather than *p*CA, in agreement with neutron diffraction studies (10).

In the O_{Glu46}–O_{pCA} bond, neutron diffraction studies (10) suggested that the H atom is located at distances of 1.21 Å and 1.37 Å from Glu46 and *p*CA, respectively. In contrast, the present QM/MM studies suggested that the H atom is at a distance of 1.00 Å and 1.58 Å from Glu46 and *p*CA, respectively, irrespective of the consistency in the O_{Glu46}–O_{pCA} distance. In agreement with the present QM/MM result, deprotonated *p*CA and protonated Glu46 were observed in experimental studies (*3-5,*

22). Hence, to explain the short O_{Glu46}–O_{pCA} distance of 2.57 Å, it is not prerequisite to locate an H atom near the midpoint between O_{Glu46} and O_{pCA}.

The O_{Tyr42}–O_{pCA} bond possessed an asymmetric double-well potential (17) (Figure 2a), which agrees with the conclusion from the neutron diffraction study that the short H bond between Tyr42 and *p*CA was not a LBHB (10). The energy value at 0.96 Å from O_{Tyr42}, which is the corresponding H atom position in neutron diffraction studies (10), was only ~0.9 kcal/mol higher than that at 1.01 Å for the energy minimum in the QM/MM geometry, i.e., essentially the same (Figure 2a). The energy near the O_{pCA} moiety is higher than that in the O_{Tyr42} moiety (i.e., near energy minimum), which indicates that the pK_a of Tyr42 is higher than that of *p*CA. The O_{Glu46}–O_{pCA} bond also possessed an asymmetric double-well potential, also corresponding to a classical H bond (Figure 2b). There is no energy minimum near 1.21 Å from O_{Glu46}, and the energy is ~5 kcal/mol higher than that at 1.00 Å. Thus, we did not essentially observe differences in the potential-energy profile between O_{Tyr42}–O_{pCA} and O_{Glu46}–O_{pCA}. The asymmetric potential curves obtained for Tyr42–*p*CA and Glu46–*p*CA cannot be classified to those of LBHB (18).

We also analyzed the potential-energy profile by altering the O_{Glu46}–O_{pCA} distance. An O_{Glu46}–O_{pCA} distance where the H atom is located nearly at the midpoint of O_{Glu46}–O_{pCA} was obtained at ~2.3 Å (Figure 3). The potential-energy curve with O_{Glu46}–O_{pCA} = 2.32 Å resembles that of a single-well potential, but is not yet completely symmetric (i.e., the minimum is not at the center of the O_{Glu46}–O_{pCA} bond (18)) due to the originally larger pK_a value of Glu46 with respect to *p*CA in the PYP environment (3-5). Obviously, even in this case the O_{Glu46}–O_{pCA} bond is unlikely to satisfy the condition of LBHB. As suggested by Schutz and Warshel (18), identification of LBHB with a single minimum potential can be valid only if the minimum is at the center of the O_{Glu46}–O_{pCA} bond. More importantly, the energy minimum with O_{Glu46}–O_{pCA} = 2.32 Å is obviously energetically higher than that with O_{Glu46}–O_{pCA} = 2.57 Å. Hence, the common case of asymmetric single minimum H bond is not LBHB but a standard H bond where the pK_a difference between the donor and acceptor moieties is large; this has been already demonstrated by Schultz and Warshel (18). Thus, unless the pK_a values of the donor and acceptor moieties are already similar in the original geometry, the resulting energy minimum of the bond is affected more by the moiety whose pK_a value is lower, and as a result, the bond becomes energetically unstable before decreasing the O_{Glu46}–O_{pCA} distance.

pK_a values of Glu46 and *p*CA in PYP. Schutz and Warshel (18) concluded that the determination of the pK_a values for the donor and acceptor moieties is the clearest way of examining the LBHB proposal. From the unusual H atom position of the Glu46-*p*CA bond, the authors of Ref. (10) speculated that the pK_a values of Glu46 and *p*CA would be similar in the PYP ground state, contributing to the LBHB formation. In contrast, we calculated the pK_a values of Glu46 and *p*CA to be

8.6 and 5.4, respectively (Table 2) (*19*); these values are in agreement with a number of experimental studies that attributed the two pK_a values near ~9 and ~6 to those of Glu46 and *p*CA, respectively (3-5). Hence, there is little basis of emphasizing the equal pK_a values for Glu46 and *p*CA in the PYP ground state.

Protonation state of Arg52. One of the backgrounds to propose the Glu46–*p*CA bond as LBHB was the interpretation of Arg52 as being deprotonated in the neutron diffraction analysis (10). Arg52 has two H-bond partners, the backbone carbonyl O atoms of Thr50 and Tyr98 (Figure 4). The $N_{\eta 1}$ atom of Arg52 appears to have only a single nuclear density toward Thr50 according to Ref. (10). In contrast, Arg52 was proposed to be protonated in a previous X-ray diffraction study (7) because the residue is exposed to the solvent. Hence, even if the basic residues are in the neighborhood of Arg52, the electrostatic influence on Arg52 will be considerably shielded by bulk water.

a) electrostatics. The calculated pK_a value of Arg52 was 13.7, suggesting that Arg52 is undoubtedly protonated (Table 2). Lys60 is the residue that most significantly contributes to Arg52 deprotonation, but decreases the Arg52 pK_a value by only 0.4 (*19*), which is located more than 7 Å away from Arg52. Thus, we could not find any reasonable mechanism to favor deprotonation of Arg52 based on the geometry of the PYP crystal structure (10).

b) QM/MM. We also independently performed QM/MM calculations to carefully evaluate the protonation state of Arg52. When Arg52 was deprotonated as suggested in Ref. (10), the N_{Arg52}–O_{Thr50} distance of 2.93 Å in the original crystal structure was significantly increased to 3.78 Å, which does not form H bonds (Table 3) (*19*). The corresponding distance with protonated Arg52 was 3.14 Å. The N_{Arg52}–O_{Tyr98} distance of 2.98 Å in the original crystal structure remains unchanged in the QM/MM geometry optimized in the presence of protonated Arg52, whereas it was increased to 3.31 Å in the presence of deprotonated Arg52 (Table 3). The resulting RMSD of the optimized side chain heavy atoms of Arg52 relative to the crystal structure (PDB: 2ZOH) was 0.143 Å for protonated Arg52 but 0.349 Å for deprotonated Arg52, the latter being above the uncertainty at a resolution of ~1 Å. Clearly, the positive charge of protonated Arg52 is prerequisite to the strong interaction with the two backbone carbonyl groups, as deprotonated Arg52 is too weak to maintain the H-bond distance in particular with Thr50.

The two independent results of the pK_a and QM/MM calculations demonstrated that Arg52 is highly likely to be protonated. The LBHB concept that was employed to explain the stabilization of the ionized *p*CA in the absence of protonated Arg52 may need to be reevaluated. If protonated Arg52 is the case for the PYP ground state, employment of the LBHB concept will be less relevant.

δ_H for PYP. Using the QM/MM optimized geometry, we calculated the δ_H values for the O_{Glu46}–O_{pCA} and O_{Tyr42}–O_{pCA} bonds and found them to be 14.5 and 14.6 ppm (PDB, 2ZOH (*10*)) or 14.6 and 14.0 (PDB, 1OTB (*9*)), respectively (*20*); these values differ by 0.6-0.7 and 0.3-0.9 ppm from the experimental values of 15.2 and 13.7 ppm (*11*), respectively (Table 4). The discrepancy may also reflect the distribution of the H-bond lengths, even in these high-resolution crystal structures of PYP at resolutions of ~1 Å (reviewed in Ref. (*9*)).

On the other hand, the H atom positions (lengths) of the neutron diffraction study (*10*) yielded δ_H values of 19.7 and 10.2 ppm for O_{Glu46}–O_{pCA} and O_{Tyr42}–O_{pCA}, respectively (Table 4) (*20*). The obtained δ_H value of 19.7 ppm satisfies the criterion of LBHB proposed by Frey (δ_H of 17 to 19 ppm (*15*)). The fact that the H atom position near the midpoint of O_{Glu46}–O_{pCA} resulted in δ_H = 19.7 ppm for typical LBHB (*15*) is also a clear validation of the criterion proposed by Schutz and Warshel (*18*), i.e., the minimum of the potential-energy curve in an LBHB is at the center of the O_{Glu46}–O_{pCA} bond. However, δ_H = 19.7 ppm is obviously larger than the value of 15.2 ppm obtained in NMR studies (*11*). Furthermore, the presence of the "LBHB" character in O_{Glu46}–O_{pCA} appears, in turn, to significantly decrease δ_H for O_{Tyr42}–O_{pCA} to 10.2 ppm relative to the experimental value of 13.7 ppm (Table 4). Hence, the H atom positions obtained in the neutron diffraction study (*10*) results in overestimation of the chemical shift for O_{Glu46}–O_{pCA} and underestimation of that for O_{Tyr42}–O_{pCA}.

3. Conclusion

We reproduced the two short H-bond donor–acceptor distances for Tyr42–*p*CA (2.50 Å) and Glu46–*p*CA (2.57 Å) using the QM/MM calculations for the entire PYP. Tyr42 was protonated and *p*CA was ionized. In the O_{Glu46}–O_{pCA} bond, a H atom was located at ~1 Å from Glu46, and Glu46 was obviously protonated in the presence of deprotonated *p*CA (Table 1). The potential-energy profile of the O_{Glu46}–O_{pCA} bond revealed an asymmetric double-well potential (17). Furthermore, there was no energy minimum near the midpoint of the O_{Glu46}–O_{pCA} bond or the *p*CA moiety (Figure 2b). We also calculated the pK_a values for Glu46 and *p*CA to be 8.6 and 5.4, respectively (Table 2). Obviously, the negative charge on ionized *p*CA is not a bare charge but already stabilized by the short H bond of Tyr42. The calculated pK_a value of Arg52 on the protein surface was 13.7, suggesting that Arg52 was protonated in the PYP ground state as in previous structural studies (7). QM/MM geometries also resulted in a significantly smaller RMSD with protonated Arg52 (0.14 Å) than with deprotonated Arg52 (0.35 Å) relative to the original crystal structure (Table 3). The NMR chemical shift of 15.2 ppm for the O_{Glu46}–O_{pCA} bond is smaller than that for single-well H bonds (20-22 ppm (*15*)) or even for LBHB (17-19 ppm (*15*)).

Although one might possibly argue that comparison between the solution NMR data (*11*) and the calculated δ_H values (solid-state crystal structure) is not relevant, one also should not ignore the resulting sufficiently high correlation between the two properties (*20*). Even if the uncertainty of maximum ~1 ppm between measured and calculated δ_H values (*20*) is considered, δ_H = 14.5–14.6 ppm obtained in the H atom position of the QM/MM geometries is reasonably closer to δ_H = 15.2 ppm obtained in the NMR studies than δ_H = 19.7 ppm obtained in the H atom position of the neutron diffraction studies (Table 4).

Thus, there is no necessity to stabilize ionic *p*CA by LBHB. The short H-bond distance of 2.57 Å for the O_{Glu46}–O_{pCA} bond can be simply explained by electrostatic interaction without invoking the LBHB concept.

References

1. Sprenger, W. W., Hoff, W. D., Armitage, J. P., and Hellingwerf, K. J. (1993) The eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein, *J Bacteriol 175*, 3096-104.
2. Baca, M., Borgstahl, G. E., Boissinot, M., Burke, P. M., Williams, D. R., Slater, K. A., and Getzoff, E. D. (1994) Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry, *Biochemistry 33*, 14369-14377.
3. Kim, M., Mathies, R. A., Hoff, W. D., and Hellingwerf, K. J. (1995) Resonance Raman evidence that the thioester-linked 4-hydroxycinnamyl chromophore of photoactive yellow protein is deprotonated, *Biochemistry 34*, 12669-12672.
4. Xie, A., Hoff, W. D., Kroon, A. R., and Hellingwerf, K. J. (1996) Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein, *Biochemistry 35*, 14671-14678.
5. Demchuk, E., Genick, U. K., Woo, T. T., Getzoff, E. D., and Bashford, D. (2000) Protonation states and pH titration in the photocycle of photoactive yellow protein, *Biochemistry 39*, 1100-13.
6. Pan, D., Philip, A., Hoff, W. D., and Mathies, R. A. (2004) Time-resolved resonance raman structural studies of the pB' intermediate in the photocycle of photoactive yellow protein, *Biophys J 86*, 2374-2382.
7. Borgstahl, G. E., Williams, D. R., and Getzoff, E. D. (1995) 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore, *Biochemistry 34*, 6278-6287.

8. Getzoff, E. D., Gutwin, K. N., and Genick, U. K. (2003) Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation, *Nat Struct Biol 10*, 663-668.
9. Anderson, S., Crosson, S., and Moffat, K. (2004) Short hydrogen bonds in photoactive yellow protein, *Acta Crystallogr D Biol Crystallogr 60*, 1008-1016.
10. Yamaguchi, S., Kamikubo, H., Kurihara, K., Kuroki, R., Niimura, N., Shimizu, N., Yamazaki, Y., and Kataoka, M. (2009) Low-barrier hydrogen bond in photoactive yellow protein, *Proc Natl Acad Sci U S A 106*, 440-444.
11. Sigala, P. A., Tsuchida, M. A., and Herschlag, D. (2009) Hydrogen bond dynamics in the active site of photoactive yellow protein, *Proc Natl Acad Sci U S A 106*, 9232-7.
12. Cleland, W. W., and Kreevoy, M. M. (1994) Low-barrier hydrogen bonds and enzymic catalysis, *Science 264*, 1887-1890.
13. Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases, *Science 264*, 1927-30.
14. Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford.
15. Frey, P. A. (2006) *in Isotope Effects in Chemistry and Biology (Kohen, A. and Limbach, H.-H., Eds.)*, pp 975-993, CRC press, Boca Raton, FL.
16. Frey, P. A. (2001) Strong hydrogen bonding in molecules and enzymatic complexes, *Magn. Reson. Chem. 39*, S190-S198.
17. Perrin, C. L., and Nielson, J. B. (1997) "Strong" hydrogen bonds in chemistry and biology, *Annu Rev Phys Chem 48*, 511-544.
18. Schutz, C. N., and Warshel, A. (2004) The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp... His pair in serine proteases, *Proteins 55*, 711-723.
19. Saito, K., and Ishikita, H. (2012) Energetics of short hydrogen bonds in photoactive yellow protein, *Proc Natl Acad Sci U S A 109*, 167-172.
20. Saito, K., and Ishikita, H. (2012) H atom positions and nuclear magnetic resonance chemical shifts of short H bonds in photoactive yellow protein, *Biochemistry 51*, 1171-7.
21. Saito, K., Ishida, T., Sugiura, M., Kawakami, K., Umena, Y., Kamiya, N., Shen, J.-R., and Ishikita, H. (2011) Distribution of the cationic state over the chlorophyll pair of photosystem II reaction center, *J. Am. Chem. Soc. 133*, 14379-14388.
22. Kandori, H., Iwata, T., Hendriks, J., Maeda, A., and Hellingwerf, K. J. (2000) Water structural changes involved in the activation process of photoactive yellow protein, *Biochemistry 39*, 7902-9.
23. Kroon, A. R., Hoff, W. D., Fennema, H. P., Gijzen, J., Koomen, G. J., Verhoeven, J. W., Crielaard, W., and Hellingwerf, K. J. (1996) Spectral tuning, fluorescence, and photoactivity

in hybrids of photoactive yellow protein, reconstituted with native or modified chromophores, *J Biol Chem 271*, 31949-31956.

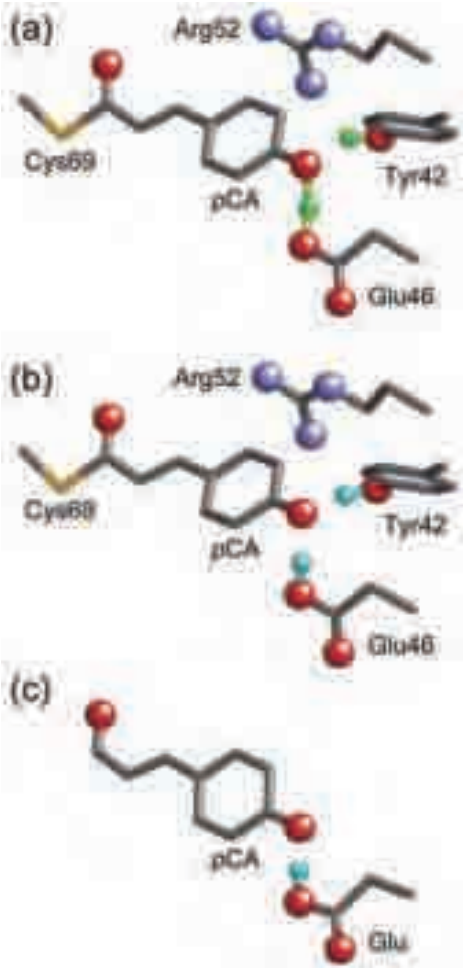


Figure 1. Geometry of the photoactive site in PYP. Only the H atom position of the H bonds between Tyr42 and *p*CA and between Glu46 and *p*CA are shown by the cyan spheres. (a) Neutron diffraction analysis (PDB; 2ZOI). (b) QM/MM optimized structure based on the X-ray crystal structure (PDB; 2ZOH). (c) The Glu– *p*CA model system.

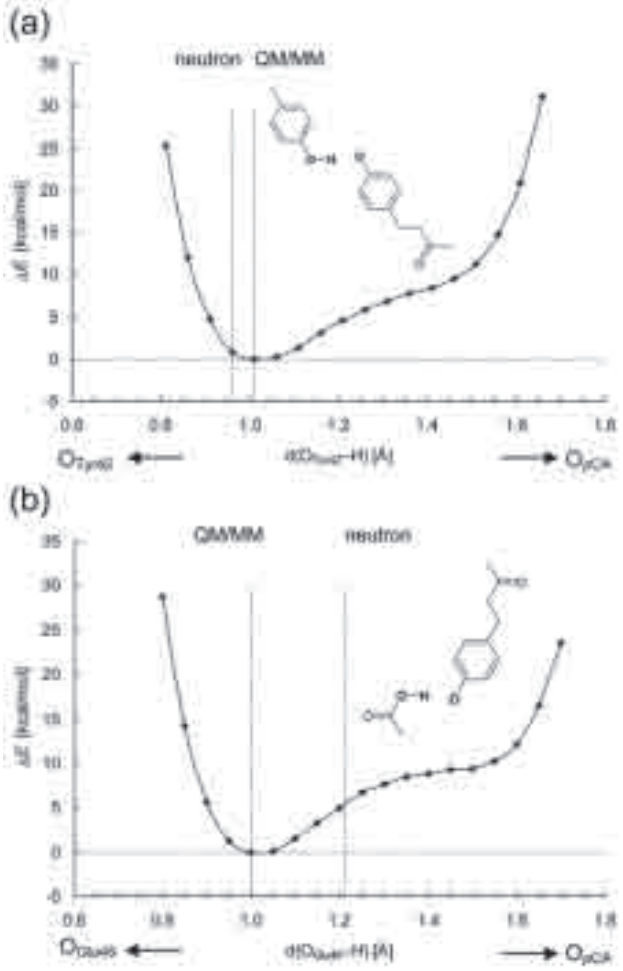


Figure 2. Energy profiles along the proton transfer coordinate for H-bond donor–acceptor pairs (a) Tyr42– *p*CA, (b) Glu46– *p*CA in the PYP protein environment, and (c) Glu– *p*CA in the model system (Figure 1c).

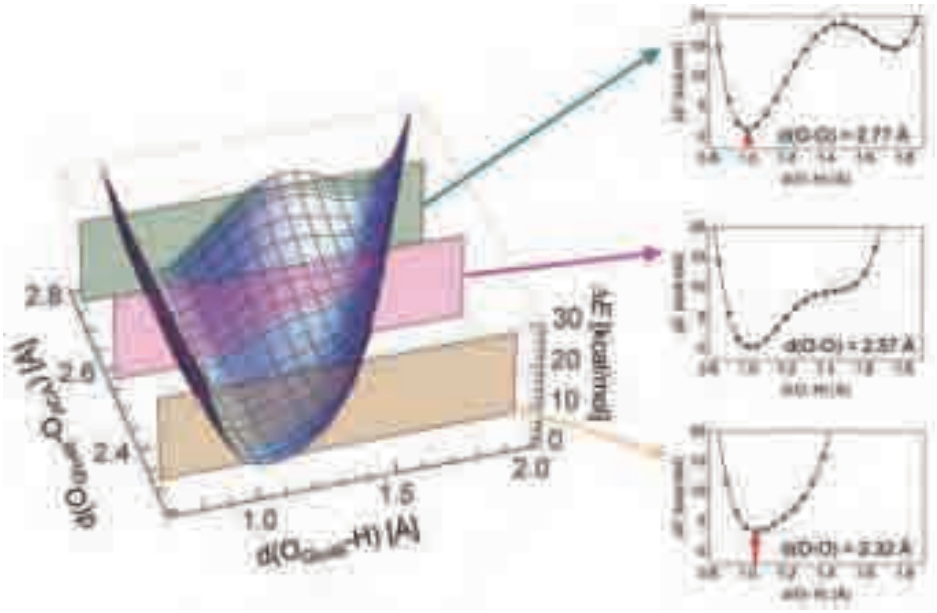


Figure 3. Dependence of the potential-energy profiles at O_{Glu46}–O_{pCA} = 2.32, 2.57, and 2.77 Å. The red arrow indicates the energy difference from the energy minimum obtained at O_{Glu46}–O_{pCA} of 2.57 Å.

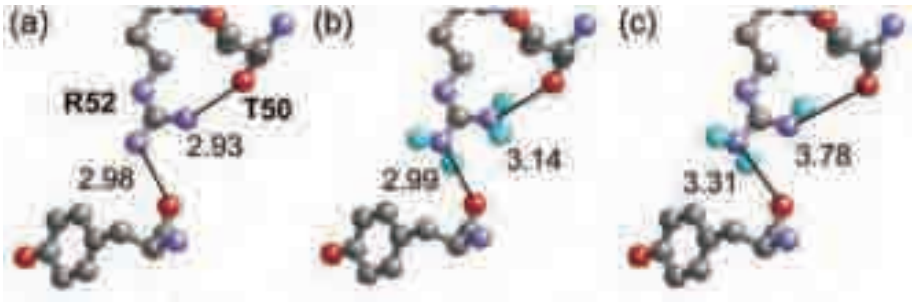


Figure 4. Influence of the Arg52 protonation state on the H-bond geometry. H atoms are shown only for the QM/MM optimized geometry (cyan spheres). (a) Crystal structure (PDB: 2ZOH). QM/MM optimized geometry with (b) protonated Arg52 and (c) deprotonated Arg52.

Table 1. H-bond distances in optimized geometries in the PYP protein environment and model systems (in Å). The H-atom positions between the H-bond donor and acceptor atoms are indicated in bold. [a] PDB: 2ZOI (10). [b] PDB: 2ZOH (10). [c] –, not applicable.

		Crystal		**QM/MM (protein)**	**Models (vacuum)**	
	Structure	neutron [a]	X-ray [b]	wild type	Glu–*p*CA	Tyr–*p*CA
Tyr42–*p*CA	O_{Tyr}–O_{pCA}	2.52	2.53	2.50	– [c]	2.57
	O_{Tyr}–**H**	0.96	– [c]	1.01	– [c]	1.03
	H–O_{pCA}	1.65	– [c]	1.50	– [c]	1.55
Glu46–*p*CA	O_{Glu}–O_{pCA}	2.57	2.57	2.57	2.52	– [c]
	O_{Glu}–**H**	1.21	– [c]	1.00	1.05	– [c]
	H–O_{pCA}	1.37	– [c]	1.58	1.47	– [c]

Table 2. Calculated pK_a values. [a] See Ref. (*23*). [b] The protein dielectric volume (i.e., the space obtained by merging the volumes of the van der Waals spheres of all protein atoms) that decreases the availability of solvation energy.

		WT			**R52A**	
		Glu46	***p*CA**	**Arg52**	**Glu46**	***p*CA**
pK_a	protein	8.6	5.4	13.7	8.1	5.9
	water (reference)	4.4	8.8 [a]	12.0	4.4	8.8 [a]
pK_a shift (water → protein)	total	4.2	–3.4	1.7	3.7	–2.9
	(i) van der Waals volume [b]	8.1	4.7	–4.7	8.0	4.4
	(ii) atomic charge	–3.9	–8.1	6.4	–4.3	–7.3

Table 3 Comparison of the H-bond donor–acceptor distances of Arg52 between the crystal structure (10) and the QM/MM optimized geometry with protonated /deprotonated Arg52 (in Å). O_{CO} = backbone carbonyl O atom. RMSD = root-mean-square deviation of the optimized side chain heavy atoms of Arg52 with respect to those of the crystal structure. –, not applicable.

	Crystal	**QM/MM (protein)**	
	X-ray	protonated	deprotonated
Arg52-Thr50 ($N_{\eta 1}$-O_{CO})	2.93	3.14	3.78
Arg52-Tyr98 ($N_{\eta 2}$-O_{CO})	2.98	2.99	3.31
RMSD	–	0.143	0.349

Table 4. Experimental (*11*) and calculated values of δ_H for short H bonds in PYP in ppm. Distances (*19*) are in Å. The neutron diffraction geometry (2ZOI) was optimized independently (i) with fixing the heavy atom positions and the two H atoms in O_{Tyr42}–O_{pCA} and O_{Glu46}–O_{pCA} or (ii) with fixing only the O_{Tyr42}–H–O_{pCA} and O_{Glu46}–H–O_{pCA} lengths. The two cases resulted in the identical chemical shifts.

	Exp. δ_H	Calc. δ_H		
Geometry (PDB)		QM/MM (2ZOH)	QM/MM (1OTB)	Neutron (2ZOI)
Glu46	15.2	14.5	14.6	19.7
(O_{Glu}–O_{pCA})		(2.57)	(2.58)	(2.57)
(O_{Glu}–H)		(1.00)	(1.02)	(1.21)
(H–O_{pCA})		(1.58)	(1.57)	(1.37)
Tyr42	13.7	14.6	14.0	10.2
O_{Tyr}–O_{pCA}		(2.50)	(2.51)	(2.52)
O_{Tyr}–H		(1.01)	(1.01)	(0.96)
H–O_{pCA}		(1.50)	(1.51)	(1.65)

18

Disulfide bond formation networks and chaperone interactions that are subject to feedback regulation

Adam Benham
School of Biological and Biomedical Sciences, Durham University, UK

Summary: In Biology, emergent phenomena can be seen all around us and are often the result of living organisms interacting with their physical environment. However, emergence and feedback can also operate at much lower scales (e.g. micrometers to nanometers) within a living cell. Here, we will contemplate the extent to which biochemical reactions and metabolism can be considered emergent, and how proteins can fold and come together to form different complexes using disulfide bridges. Disulfide bridges are the result of two SH (sulfur and hydrogen) chemical groups coming together to form an S-S bond. The sophisticated way in which this is controlled, by proteins called chaperones, and is subjected to feedback by the cell will be discussed.

1. Introduction:

Emergence and feedback are important concepts in Biology. George Lewes, the philosopher and critic who first introduced the idea of emergence to Victorian Britain, was himself very interested in biological phenomena, particularly the organization of the nervous system. Lewes wrote in *Problems of Life and Mind* that emergent effects are not additive and cannot be predicted from the component parts or "decomposed" into the original components. These emergent phenomena are visible all around us in the natural world (1). Classic examples range from the variety and complexity of termite and ant nests to the swarming behaviors of birds and fish (Figure 1).



Figure 1: *(A) a termite mound, (B) a cast of an ants nest, (C) schooling fish in the Maldives and (D) a flock of birds in Silkeborg, Denmark illustrate how animal behavior can give rise to complexity in physical systems. Photographs courtesy of a creative commons license.*

The examples shown in Figure 1 illustrate emergence on a relatively large scale, where individual organisms combine to create higher order structures that are visible to the naked eye. However, emergence occurs at all scales and its consequences can be observed at the molecular level. In Biochemistry, emergence can result from simple chemical reactions. For example, when sodium hydroxide (a base) and hydrochloric acid react together they produce water and salt ($NaOH + HCl \rightarrow NaCl + H_2O$). The products of this reaction can be considered "emergent" because they are not the sum of the reaction as such, but have completely different properties from the original reactants. The salt solution produced in the reaction described above is neutral and could even be drunk. But drinking sodium hydroxide and hydrochloric acid is extremely dangerous and could be fatal!

In a cell or tissue of a biological organism, a wide range of biochemical reactions like the reaction between an acid and a base occur all the time. It is the temporal and spatial organization and control of these reactions that ultimately gives rise to a functional and unique organism (2). Feedback is particularly important in controlling the biochemical reactions within cells and tissues in a living being. Feedback mechanisms are important because they prevent reactions from running out of control and make optimal use of scarce resources and energy. Frequently, the product of a reaction governs feedback inhibition. A classical example to illustrate biochemical feedback is the biosynthesis of the amino acid isoleucine in bacteria (3). All living organisms require amino acids because they are the building blocks for proteins. Bacteria make isoleucine from another amino acid, called threonine, in a stepwise process illustrated schematically in Figure 2.



Figure 2: *Threonine (compound A) is converted to α-Ketobutyrate (compound B) by threonine deaminase. A series of enzymes converts α-Ketobutyrate to isoleucine (compound Z). Isoleucine can feedback-inhibit threonine deaminase and thus stop its own production. Image taken from Stryers "Biochemistry".*

The reactions shown in Figure 2 do not occur spontaneously but require enzymes to catalyse them. The first enzyme in the pathway, threonine deaminase, is inhibited by isoleucine. So when the level of isoleucine is high enough, its production is switched off. Isoleucine binds to a regulatory site on threonine deaminase that is set apart from the catalytic site of the enzyme and mediates its effects by reversible (allosteric) interactions. This type of feedback regulation ensures that the enzyme can be "recycled" and can resume isoleucine synthesis when levels of the product drop back down. This system of feedback seems so clever that it is hard to image that it wasn't designed. Of course, feedback inhibition of enzymes has come about through natural selection. Any bacteria with the ability to optimize their amino acid resources will have a growth advantage over their competitors. Some bacteria like *Escherichia coli* have a generation time of ~20 minutes when growing in nutrient rich conditions, so it does not take long for an advantageous mutation to become fixed in a bacterial population.

In life, not all biosynthetic reactions are as simple as those depicted in Figure 2 and more complex feedback mechanisms exist that govern a wide range of metabolic processes (4). For example, when a metabolic pathway branches, such as in the synthesis of aromatic amino acids in *Bacillus subtilis*, the first divergent steps in the pathway can be inhibited by the final products of each individual branch (Figure 3). This causes the build up of an intermediate in the pathway that can inhibit the first common step. In this way, the synthesis of one product is not blocked by the other product and the first step is only shut down if both metabolic products are present in excess.

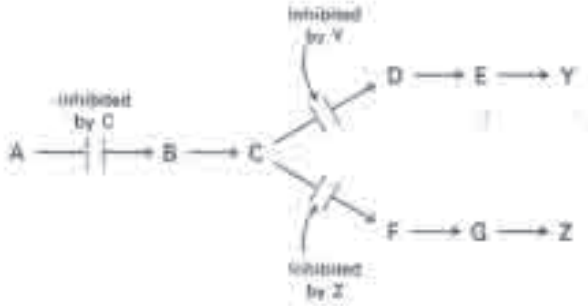


Figure 3: *Sequential feedback control in aromatic amino acid synthesis. The end products, Y (phenylalanine) and Z (tryptophan), can inhibit enzymes controlling their own synthesis after a common branchpoint. Inhibiting both branches causes the build up of intermediate product C, which can inhibit the whole pathway. For simplicity, letters have been used to denote the amino acid intermediates. Image taken from Stryers "Biochemistry".*

There are 20 commonly occurring amino acids that make up our proteins. One amino acid that my laboratory is particularly interested in is cysteine. Cysteine and methionine are the only regular amino acids that contain a sulfur (S) atom (5). In cysteine, the sulfur atom is positioned at the end of the molecule and because sulfur is highly reactive, this gives cysteine the ability to make covalent bonds, called disulfide bridges, with other sulfur containing molecules. For example, when two cysteine amino acids react together, they form oxidized cystine (Figure 4).

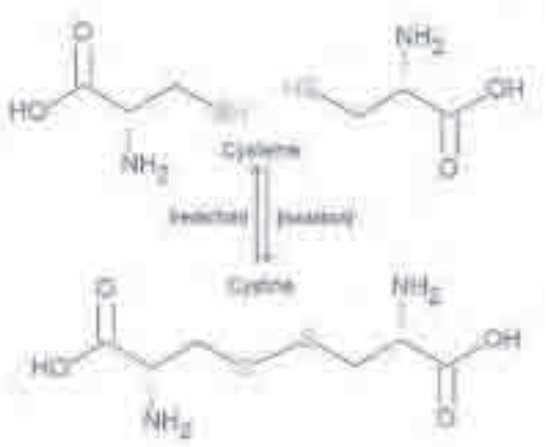


Figure 4: *Cysteine has a free sulfhydrl group (-SH) that can react with another free –SH group on a cysteine to form a cystine molecule. A free –SH group is reduced and an S-S bond is oxidized.*

Cysteine does not just exist as a stand-alone amino acid, but is incorporated into proteins during protein synthesis. Proteins with cysteine residues in them, therefore, are also able to form disulfide bridges and cysteine-containing proteins can exist in either a reduced (-SH) or oxidized (S-S) form. The disulfide bridge is a very useful molecular tool and is used by the cell in a variety of ways. The disulfide bridge can make a protein more compact and help it fold up into structures that can resist the harsh environment outside the cell (Figure 5). This is one reason why proteins with disulfide bonds are most commonly found in proteins that are secreted or located at the cell membrane (6). Disulfide bridges can also be used to connect proteins together, including two copies of the same protein (called a homodimer), two different protein types (a heterodimer) or many different proteins (oligomers).

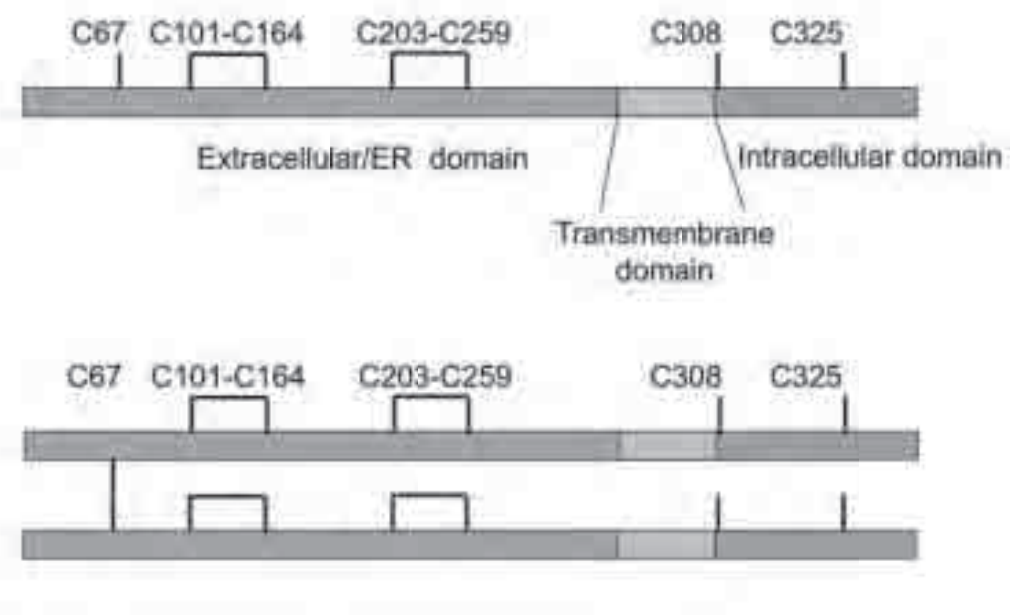


Figure 5: *Disulfide bonds can link two parts of the same protein together (an intramolecular disulfide bond e.g. between C101-C164 and C203-C259) or can link two individual proteins together (an intermolecular disulfide bond e.g. C67-C67 shown in the lower panel). C= cysteine, numbers represent the position of the cysteine amino acid in the protein chain.*

The reversibility of the disulfide bond provides the cell with a way to make complex emergent structures from more simple building blocks, and offers an opportunity to regulate these structures by selective reduction or oxidation of the disulfide bonds. Oxidation involves the loss of electrons, so proteins containing cysteine are often involved in reactions that help transport electrons around the cell. Electron transport is important to cellular wellbeing for various reasons; one of the most critical roles is in the mitochondria, where electron transport helps to establish a proton gradient across the inner mitochondrial membrane and thus drives the synthesis of ATP, the energy currency of the cell (7).

Although cysteine is a very reactive amino acid, the reduced and oxidized forms exist in equilibrium, and whether the reduced (–SH) form or the oxidized (S-S) form is favored is governed by the reduction potential of the chemical surroundings. The reduction potential is a measure of the tendency of a chemical entity to acquire electrons and become reduced and is determined by a number of factors

including pH, temperature and other chemical substances. Within a cell, in a complex organism such as *Homo sapiens*, different compartments perform different cellular functions and each of these has a slightly different reduction potential. The place where most secreted proteins are assembled in higher organisms is called the endoplasmic reticulum and this organelle has a more oxidizing potential than the surrounding cytoplasm. This means that proteins in the endoplasmic reticulum are more likely to form disulfide bonds between two cysteine residues. Left unattended, proteins in the endoplasmic reticulum could easily form inappropriate disulfide bonds and become dysfunctional. Thus disulfide bond formation is subjected to stringent quality control and regulation.

In fact, the introduction of disulfide bonds into proteins within the endoplasmic reticulum is controlled by a small set of enzymes that have the capacity to "donate" disulfide bonds to a range of client proteins and help them fold along a productive pathway. The major enzyme that controls this process is Protein Disulfide Isomerase (PDI) (8). PDI can introduce, remove or shuffle disulfide bonds in a variety of proteins (Figure 6). PDI is also classed as a molecular chaperone, which means that it can assist the folding and assembly of its protein clients.

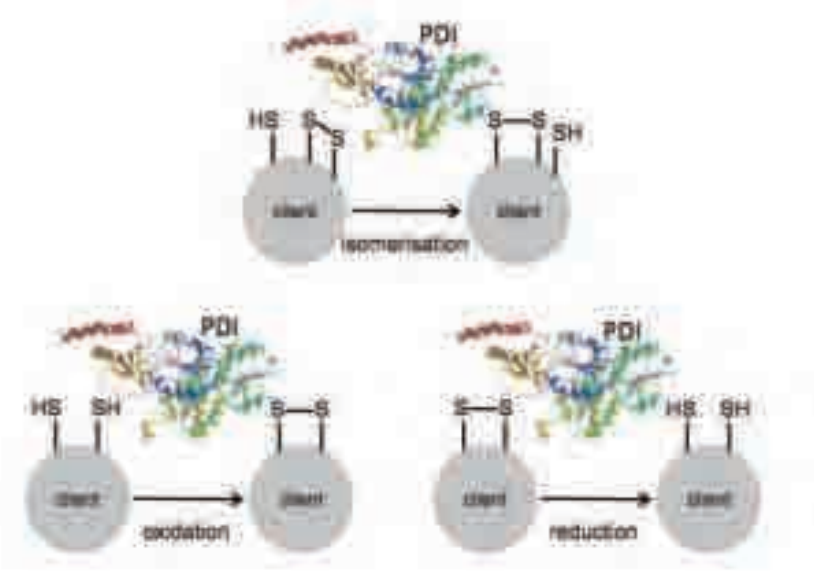


Figure 6: *PDI (coloured structure) can isomerise (rearrange), oxidize (insert) or reduce (remove) disulfide bonds in a variety of target proteins (depicted as grey circles).*

Over the course of evolution, the PDI gene has been duplicated and in humans there are at least 20 different versions of PDI (called PDI homologs) (9). Despite the fact that much is known about PDI itself, we do not know the exact function of most of the PDI family members and this is a major area of ongoing research. Some of these PDIs appear to have tissue specific or specialist roles to play in cellular function. For example, a relative of PDI that my group first described in 2005 has recently been shown to be important in controlling male fertility (10,11).

In general, when PDI oxidizes a client protein, it will become reduced, so PDI needs to be reoxidised if it is to be used more than once by the cell. To complete the oxidation cycle, a protein called

Endoplasmic Reticulum Oxidoreductase (ERO) recharges PDI. What then reoxidises ERO? The answer is molecular oxygen. A cofactor called flavin adanine dinucleotide (FAD) passes on the electrons generated by the PDI-ERO pathway to oxygen, in essence generating a disulfide bond from thin air (12). This process could have potentially dangerous consequences for the cell, because when oxygen gains electrons it forms toxic free radicals, including hydrogen peroxide, which can damage bystander proteins, lipids and nucleic acids. To prevent this sort of oxidative damage from occurring during disulfide bond formation, the cell has a number of feedback and control mechanisms in place that help to limit the generation of reactive oxygen species (13). The control mechanisms include enzymes called peroxiredoxins that can use the potentially dangerous peroxide to provide more disulfide bonds for newly synthesized proteins. The extent of these feedback and control mechanisms that regulate oxidative protein folding is only recently being realized and the opportunities for future research in this area will be discussed further in the presentation.

2. Discussion:

Biologists frequently strive to predict emergent effects from the sum of the components. This, in essence, is part of reductionist thinking and the notion that complex systems can be understood in terms of their constituent parts. A reductionist approach to experimental biology and biochemistry is very valuable because it enables us to control the system and eliminate variability. However, to understand emergent functional complexity in Biology, rather than merely to observe and describe Biological complexity, requires fresh thinking and a multidisciplinary approach. Bridging the gap is not easy. Returning to the example of proteins, we are now very adept at solving protein structures by crystallography and other techniques, and we can predict quite accurately where a disulfide bond will be located in proteins whose structures have not been solved. However, protein structures and large protein assemblies remain very difficult to predict from the primary amino acid sequence information alone. Understanding the emergent complexity and feedback regulation of disulfide bond networks in the cell is a work in progress. It is becoming increasingly apparent that the disulfide bond machinery can become dysregulated in disease processes ranging from diabetes to neurodegeneration, so understanding how to manipulate these systems is of major interest for human health.

Reference:

1. Sumpter, D. J. (2006) *Philos Trans R Soc Lond B Biol Sci* **361**, 5-22
2. Kholodenko, B. N., Hancock, J. F., and Kolch, W. (2010) *Nat Rev Mol Cell Biol* **11**, 414-426
3. Brown, K. D. (1968) *Genetics* **60**, 31-48
4. Umbarger, H. E. (1978) *Annu Rev Biochem* **47**, 532-606
5. Stipanuk, M. H., and Dominy, J. E., Jr. (2006) *Amino Acids* **30**, 251-256
6. Feige, M. J., and Hendershot, L. M. (2011) *Curr Opin Cell Biol* **23**, 167-175
7. Skulachev, V. P. (1977) *FEBS Lett* **74**, 1-9
8. Hatahet, F., and Ruddock, L. W. (2009) *Antioxid Redox Signal* **11**, 2807-2850
9. Benham, A. M. (2012) *Antioxid Redox Signal* **16**, 781-789
10. van Lith, M., Hartigan, N., Hatch, J., and Benham, A. M. (2005) *J Biol Chem* **280**, 1376-1383

11. Tokuhiro, K., Ikawa, M., Benham, A. M., and Okabe, M. (2012) *Proc Natl Acad Sci U S A* **109**, 3850-3855
12. Gross, E., Sevier, C. S., Heldman, N., Vitu, E., Bentzur, M., Kaiser, C. A., Thorpe, C., and Fass, D. (2006) *Proc Natl Acad Sci U S A* **103**, 299-304
13. Bulleid, N. J., and Ellgaard, L. (2011) *Trends Biochem Sci* **36**, 485-492

19

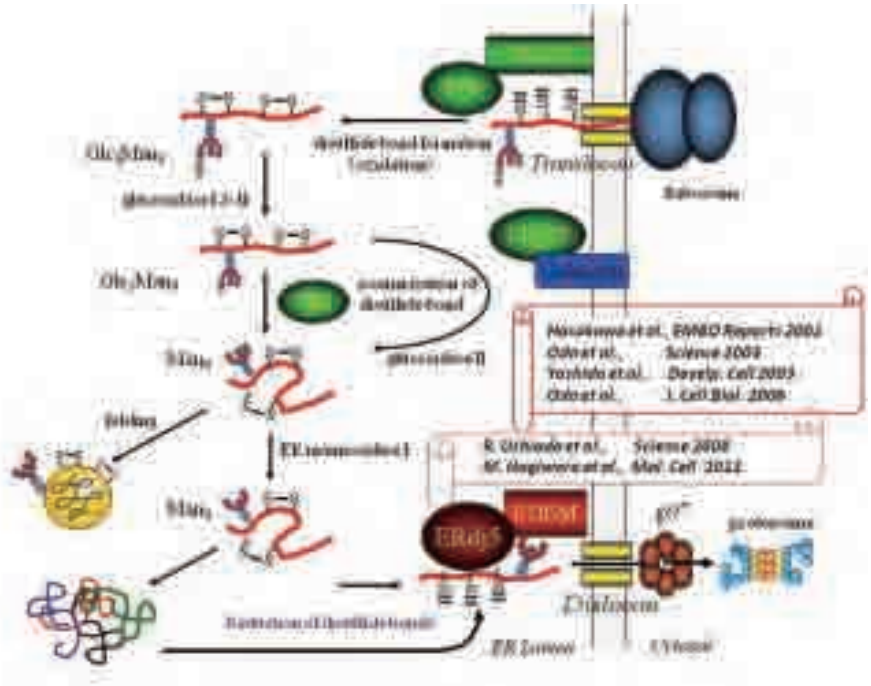
Regulation of redox homeostasis and proteostasis in the ER

Kazuhiro Nagata
Department of Molecular and Cellular Biology, Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603-8555, Japan
E-mail: nagata@cc.kyoto-su.ac.jp

Nascent polypeptides co-translationally entering the ER are co- and post-translationally modified by N-glycosylation and oxidation/isomerization of cysteine residues followed by correct folding with the aid of molecular chaperones. Oxidative environment in the ER enables ER-resident oxidoreductases including PDI family members to facilitate in making disulfide bond formation, which stabilizes the ternary or quaternary protein structures. Only properly folded proteins could reach to their final destination.

In this talk, I will focus on the involvement of the ER oxidoreductase system in both the productive folding of newly synthesized proteins and ER associated degradation (ERAD) of misfolded proteins as a quality control mechanism in the ER.

ERAD is a multistage process involving ER luminal and cytosolic events. The anterior half of ERAD is comprised of three major events; recognition and segregation of terminally misfolded proteins from folding intermediates, unfolding of the misfolded substrates by oxidoreductases that cleave the disulfide bonds to enable the substrates to be translocated through the retrotranslocation channel, and transport of substrates to be degraded to the dislocon channel. The factors required for these three critical steps form a supramolecular complex in the ER. This complex is comprised of EDEM1, a lectin-like molecule that recognizes mannose-trimming and segregates the substrates from the productive folding pathway into the degradation pathway, ERdj5, a reductase that resides in the ER and reduces the disulfides of misfolded proteins to form extended polypeptides, and BiP, an Hsp70 family molecular chaperone that recruits substrates to the dislocon channel after dissociation from the EDEM1/ERdj5 complex coupled with

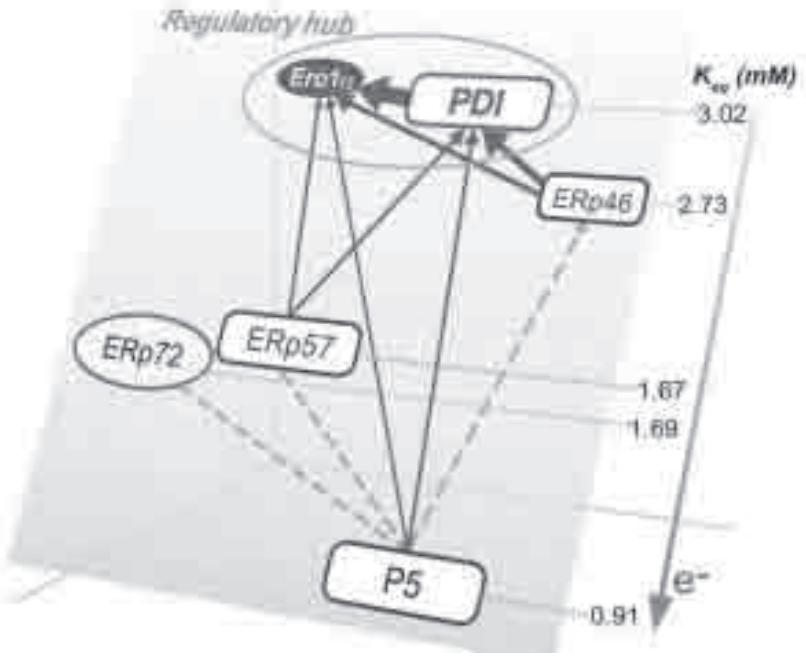


ATP hydrolysis.

We will discuss here the importance of disulfide bonds reduction in misfolded proteins for retrotranslocation through the dislocon channel by comparing the function of ERdj5 with other oxidoreductases in the ER.

While it is well established that oxidoreductases in the ER play major roles not only for the productive folding of newly synthesized proteins but also for the degradation of misfolded proteins in the ER (ERAD), the electron transfer cascade among the ER oxidoreductases as well as the networks of their interactions are poorly understood. Here, we have performed proteomic interaction analysis for more than 20 ER oxidoreductases by identifying the ERO1a-associated oxidoreductases, and have established the cascade of oxidation reaction where several oxidoreductases such as ERp46, ERp57 and P5 are sequentially oxidized by the ERO1α -PDI regulatory hub complex.

Protein homeostasis (proteostasis) and redox homeostasis in the ER are both most important among the various components contributing the maintenance of the ER homeostasis. We showed here that two homeostasis are closely correlated each other, and would like to discuss the importance of these two homeostasis in terms of the cellular reguration.



20

Application of quantum mechanical formulation to classical stochastic processes

Jun Ohkubo
Graduate School of Informatics, Kyoto University

Stochastic processes have been studied a lot in various research fields, and recently stochastic properties in small systems such as cells attract many attentions in biology and physics. My main message of this talk is that methods and concepts in quantum mechanics are sometimes useful in studies of classical stochastic processes. Although the stochastic processes have no quantum effects, there is a formal analogy between quantum mechanics and the classical stochastic processes. Interdisciplinary works based on such formal analogies will play important roles in future works. I will give an example in this talk; a scheme of counting statistics enables us to count the statistics of the number of specific transitions during a finite time interval, and a geometric phase concept in quantum mechanics is employed in order to investigate a pump current phenomenon, in which a current caused by periodic perturbations in a simple stochastic process. As a result, an intuitive and understandable picture for the pump current phenomenon is obtained.

1. Introduction.

Stochasticity and fluctuations have been observed in diverse phenomena in various research fields. The dynamics of such phenomena are formulated as stochastic processes, and hence the stochastic processes have been hot research topics in various research fields, including physics and biology. Especially, recent development of experimental techniques, such as single-molecule experiments, enables us to obtain various detailed data for small chemical systems such as cells. Studies for the small systems have clarified the fact that stochasticity and fluctuations play important roles for functions of the small systems, i.e., the biological matters.

A master equation is one of the famous schemes to describe the stochastic processes. (I will show an example of the master equation later.) The master equation consists of simultaneous differential equations. There are many methods to obtain numerical solutions for the stochastic processes; one may solve the simultaneous differential equations numerically, or the other may use Monte Carlo simulations in order to simulate the phenomena directly. However, in order to understand the stochasticity and dynamical properties of the phenomena, it would be useful to obtain analytical solutions of the master equations and to clarify the mathematical structures behind the phenomena. Hence, it is important to seek further mathematical techniques to treat the master equation, and it will have a grate impact on various research fields.

A main message of this talk is as follows: *methods and concepts in quantum mechanics are sometimes useful in studies of classical stochastic processes*. That is, despite the fact that there is no quantum effect in the classical stochastic processes, there are formal analogies between the

classical stochastic processes and quantum mechanics. If there are similar mathematical structures, it is possible to use mathematical and analytical methods developed in quantum mechanics in order to investigate the classical stochastic processes. Of course, it may not be needed to say `quantum mechanics'; we should only refer to the mathematical structures. However, there are many practical calculation schemes and analytical methods developed in quantum mechanics, and hence I here say `quantum mechanics' and employ them. Note that the meanings of the results should be treated carefully because the same mathematical structure does not mean the same physical or real phenomena.

As an example, I focus on a so-called pump current phenomenon, in which a current is caused by periodic perturbations for a system. Using a simple stochastic model, I explain a master equation, counting statistics, and an application of quantum mechanical formulations, i.e., a geometric phase, to the pump current phenomenon.

2. Particle-hopping model and a master equation

As I explained in Section 1, there is a phenomenon known as a pumping, in which a system under periodic perturbations causes a finite flux in a preferred direction. Experimentally, it has been known that a red blood cell under an oscillating electric field makes a Na^+ efflux at certain experimental conditions [1]. Questions here are *why the periodic perturbations cause a current, and what the mathematical structures behind the phenomenon are.* `Periodic' means that strength of the perturbations is changed periodically, and how does such periodic behavior make a unidirectional current? Is there a kind of feedback mechanisms behind this phenomenon?

In order to investigate the pump current phenomenon, I here use a simple particle-hopping model instead of a realistic complicated model. The particle-hopping model is shown in Figure 1 [2]. The system consists of three parts, i.e., a container and two particle reservoirs. The container can contain either zero or one particle in it. When the container is filled with one particle, the particle can escape from the container by jumping into one of the two particle reservoirs. In contrast, when the container is empty, either of the particle reservoirs can emit a new particle into the container. The particle reservoirs are very large, and we assume that the particle emission



Figure 1: Particle-hopping model. There are large particle reservoirs (left and right), and the container can contain at most one particle. If the container is filled with a particle, no particle can move into the container.

from the reservoirs and particle injection into the reservoirs have no effect on the reservoirs. k_1, k_{-1}, k_2, k_{-2} are kinetic rates or transition rates for the corresponding particle hopping.

For the simple particle-hopping model, we can construct a master equation. The master equation for the state of the container is as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} p_{\mathrm{e}} \\ p_{\mathrm{f}} \end{bmatrix} = \begin{bmatrix} -k_1 - k_{-2} & k_{-1} + k_2 \\ k_1 + k_{-2} & -k_{-1} - k_2 \end{bmatrix} \begin{bmatrix} p_{\mathrm{e}} \\ p_{\mathrm{f}} \end{bmatrix}, \tag{1}$$

where p_{e} and p_{f} is the probability with which the container is empty and filled, respectively. The master equation consists of two ordinal differential equations; one is for p_{e} and the other is for p_{f}. If the container is empty, a particle is entered from the reservoirs at rate (k_1+k_{-2}); in this case, the probability p_{e} decreases. On the other hands, if the container is filled, the particle in the container moves to the reservoirs at rate (k_{-1}+k_2), which increases the probability p_{e}. There are minus and plus contributions for the change of p_{e}, respectively. For the change of p_{f}, the similar discussions are possible, and we obtain equation (1).

As periodic perturbations, we here consider the following transition rates:

$$\begin{aligned} k_1 &= c_1 + R\cos(\omega t), & k_2 &= c_2, \\ k_{-1} &= c_{-1}, & k_{-2} &= c_{-2} + R\sin(\omega t). \end{aligned} \tag{2}$$

There are constant rates, c_1, c_{-1}, c_2, c_{-2}, and periodic perturbations. R is the amplitude of the perturbations, and ω is the angular velocity. We assume that R is small, and all transition rates are positive for any time.

By solving equation (1), the probabilities for the state of the container at a certain time are calculated. However, the aim of this talk is to evaluate a current in the system. Here, we consider a current from the container to the right reservoir: If a particle moves from the container to the right reservoir, it contributes as a positive current; a particle hopping from the right reservoir to the container gives a negative contribution. Hence, the current is defined as a subtraction of the number of [right reservoir → container] from that of [container → right reservoir]. We want to calculate the statistics of the current; that is, not only the average value, but also the deviation or higher-order statistics are calculated. However, the probabilities p_{e} and p_{f} are not enough to evaluate the current statistics. Hence, we employ here a scheme of counting statistics, which has been developed mainly in chemistry and physics [3].

3. Counting statistics

In the scheme of the counting statistics, some target transitions in the stochastic process are focused. For example, the current we consider here is related to the hopping between the container and the right reservoir. I omit the detailed explanations for the scheme (please see

references [2] or [3] for details), and I show only the consequences. In order to evaluate the statistics of the current, it is useful to calculate a generating function:

$$F(\chi, t) = \sum_{N_\mathrm{A}=-\infty}^{\infty} \mathrm{e}^{N_\mathrm{A}\chi} P(N_\mathrm{A}|t), \tag{3}$$

where $P(N_\mathrm{A}|t)$ is a probability with which there are N_A *net* transitions from the container to the right reservoir during time t. Once we obtain the generating function, all statistics of the current can be evaluated. For example, the average current is calculates as follows:

$$\langle N_\mathrm{A} \rangle = \frac{\partial}{\partial \chi} F(\chi, t)\bigg|_{\chi=1} = \sum_{N_\mathrm{A}=-\infty}^{\infty} N_\mathrm{A} P(N_\mathrm{A}|t). \tag{4}$$

The probability $P(N_\mathrm{A}|t)$ is unknown, and how should we calculate the generating function? Following the scheme of the counting statistics, it is clarified that the generating function is evaluated as

$$F(\chi, t) = f_\mathrm{e}(\chi, t) + f_\mathrm{f}(\chi, t), \tag{5}$$

where the restricted generating functions f_e and f_f are calculated as

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} f_\mathrm{e} \\ f_\mathrm{f} \end{bmatrix} = \begin{bmatrix} -k_1 - k_{-2} & k_{-1} + \mathrm{e}^{\chi} k_2 \\ k_1 + \mathrm{e}^{-\chi} k_{-2} & -k_{-1} - k_2 \end{bmatrix} \begin{bmatrix} f_\mathrm{e} \\ f_\mathrm{f} \end{bmatrix}. \tag{6}$$

Note that equation (6) is very similar to the master equation (1). The transition from the container to the right reservoir has a positive contribution to the current and it is related to the transition rate k_2, and hence the exponential factor is multiplied to k_2 in the non-diagonal part. On the other hand, the transition from the right reservoir to the container, which is related to k_{-2}, has the negative contribution and hence the inverse factor is multiplied. As exemplified here, it is easy to obtain the time-evolution equations for the restricted generating functions from the original master equation. The scheme of the counting statistics has been studied a lot; various methods in the counting statistics will be useful in diverse research fields.

4. Geometric phase discussions

It is possible to evaluate the statistics of the current numerically [4], and the numerical evaluations and Monte Carlo simulations show that actually there are a net current from the container to the right reservoir. In order to grasp the pump current phenomenon more intuitively, a mathematical analysis will be important. For the analysis, a quantum mechanical formulation,

i.e., a geometric phase, is available and useful. The geometric phase is used to study various phenomena in quantum mechanics, such as Aharonov-Bohm effects. Why is the method in quantum mechanics applicable to the classical stochastic process, i.e., the pump current? The answer is that there is a formal analogy between the time-evolution equation for the restricted generating functions (6) and a Schrödinger equation in quantum mechanics.

We write the restricted generating functions as

$$|\psi\rangle = \begin{bmatrix} f_{\mathrm{e}} \\ f_{\mathrm{f}} \end{bmatrix}, \tag{7}$$

and the time-evolution operator in equation (6) as follows:

$$-H = \begin{bmatrix} -k_1 - k_{-2} & k_{-1} + \mathrm{e}^{\chi} k_2 \\ k_1 + \mathrm{e}^{-\chi} k_{-2} & -k_{-1} - k_2 \end{bmatrix}. \tag{8}$$

Then, the time-evolution equation (6) is rewritten as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -H|\psi\rangle. \tag{9}$$

Equation (9) is similar to an imaginary-time Schrödinger equation. When the Hamiltonian, *H*, has a kind of periodicity, it is known that the geometric phase discussion is sometimes useful. Actually, it has been shown that the geometric phase discussion is applicable to the classical stochastic processes [2,5,6]. From these studies, it has been clarified that a cumulant generating function for the current, which is a kind of the generating function similar to equation (3), is naturally split into two parts:

(Cumulant generating function)
= (Contribution from the dynamical phase)
+ (Contribution from the geometric phase).

Here, I do not explain the details of the analysis, but the important point is that there is a mathematically natural way of the division of the cumulant generating function. In addition, there is a physical and realistic meaning for the division. That is, the current is split into two parts corresponding to the division of the cumulant generating function:

(Current) = (Current from difference of particle densities)
+ (Current from periodic perturbations).

The first term corresponds to the contribution from dynamical phase. If there is a difference between two particle reservoirs (chemical potential difference), a current from the dense reservoir to the sparse reservoir should occur. When $c_1 = c_2 = c_{-1} = c_{-2}$, there is no average current

caused by the difference of the particle densities, and its fluctuation can be analyzed from the dynamical phase. On the other hand, the pump current, which is caused by the periodic perturbations, corresponds to the contribution from the geometric phase. From the mathematical structures of the geometric phase, it is immediately possible to obtain various information for the pump current, such as conditions to make the pump current and the direction of the pump current, and so on.

5. Concluding remarks

In this talk, I showed an example in which a method developed in quantum mechanics is applied to an analysis of a classical stochastic process. Such interdisciplinary trials will be important in future works, because many phenomena in various research fields have similar mathematical structures and it is possible to employ methods and concepts developed in a specific research field to the other research fields. Actually, it is a common situation that a practical method needed to one's study has been already developed in other research fields. In order to tackle dynamics, emergence, and feedback problems, I believe that such interdisciplinary trials help many researchers in various research fields.

Reference:

[1] D.S. Liu, R.D. Astumian, and T.Y. Tsong, J. Biol. Chem. 265, 7260 (1990).
[2] N.A. Sinitsyn and I. Nemenman, Europhys. Lett. 77, 58001 (2007).
[3] I. Gopich and A. Szabo, J. Chem. Phys. 122, 014707 (2005); ibid. 124, 154712 (2006).
[4] J. Ohkubo and T. Eggel, J. Stat. Mech. P06013 (2010).
[5] J. Ohkubo, J. Stat. Mech. P02011 (2008); J. Chem. Phys. 129, 205102 (2008).
[6] N.A. Sinitsyn, J. Phys. A: Math. Theor. 42, 193001 (2009).

21

Emergent Relations: meanings and materiality in human-environmental interaction

Veronica Strang
Durham University

Summary This research explores how human sensory and cognitive interaction with the material properties of things generates ways of conceptualising objects and processes and encoding meaning in them. For example, people use water 'to think with' – to imagine flow, movement and change over time. Thus the material properties of water have agentive power, not only in affecting how people can engage with it physically, but also in shaping the ideas and meanings they attach to it. This reciprocal linkage between materiality and mind produces a co-evolving human-environmental relationship in which the properties of things and the behavioural properties of persons maintain spatial and temporal consistencies and thus a degree of order in complex social and ecological systems – up to a point. But things and persons also undergo continual change and development, opening the door to 'emergent' events which may be radically destabilising.

1. Introduction - Fluid Relations

We are all familiar with the material properties of water: its fluidity, its transformative capacities, its conductivity and its connectivity. The molecular structure of water is such that these properties pertain at every level: water is vital in human, animal and plant bodies, irrigating cells, enabling the circulation of nutrients, and carrying away waste. It is similarly essential to all ecosystemic processes, whether local, regional, continental or planetary in scale. Water materially connects the tiniest microbe with world hydrological systems (Helmreich 2009, Margulis and McMenamin 1992, McMenamin and McMenamin 1994, Vernadsky 1986). This connection is both spatial and temporal: the fluid movements of water may be as fleeting as the tiny capillary shifts through which water crosses a plant membrane, or as literally glacial as the accumulation of meltwater in vast underground aquifers.

Water is therefore ideal medium for considering material and ideational flows between temporal and spatial scales. In particular, it allows us to consider how the properties of things are omnipresent in and active upon engagements between things and persons, and implicit in the 'emergent' events that arise from these interactions. Further, it underlines a reality that matter is not ontologically distinct: persons are bio-cultural beings, and human-environmental relationships are composed similarly of collective interactions between material and social processes. Thus Strathern describes socially

expanded ‘hybrid’ networks of relations (1996) which Knappett reminds us are both spatially and temporally processual (2007: 22).

Drawing attention to Latour’s work on assemblages of things and persons (1995, 2005) and Deleuze’s representations of life as a ‘pulse’ or ‘movement’ in which ‘all entities constantly differentiate or transform’ (1994), Harvey highlights the utility of Actor Network Theory (ANT) in expressing the fluidity of relationships between humankind and non-human species and things, and the importance of focusing not on the people or things themselves, but on the interstitial spaces between (Harvey 2012: in press, see also Harvey and Knox 2010). Ingold takes a similarly Deleuzian view to argue that, rather than separating things and people, or earth and sky, we need to pay attention to ‘the fluxes of wind and weather’ (2007a). Thus we are presented with complex assemblages of things and persons and ideas that are made coherent or ‘sympathetic’ by the relationships between them.

> What is an assemblage? It is a multiplicity which is made up of many heterogeneous terms and which establishes liaisons, relations between them. (Deleuze and Parnet, 2006: 69)

This coheres with a theoretical shift to less anthropocentric visions of human-environmental interactions. A more ethical and egalitarian form of relationality acknowledges the agency of non-human species and things (Haraway 2008). In this regard Bennett’s work on ‘vibrant matter’ and ‘vital materiality’ is particularly useful in highlighting the dynamism of interactions between human and non-human beings and material objects (Bennett 2010). Thus anthropological theory has moved towards increasingly fluid and permeable visions of human-environmental relations. Things and persons mix in complex and constantly shifting relationships, and notions of personhood are highly contingent (Verdery and Humphrey 2004). Identities flow across boundaries; things and places become invested with personhood and incorporated into social life (Henare *et al* 2007), and ‘previously taken-for-granted notions of ‘persons’, ‘things’ and ‘relations’ are now thoroughly destabilized’ (Busse and Strang 2012: xv). Current writing on relational materiality similarly highlights the ontological fluidity in which experiences, practices, technologies and representations undergo continual renegotiation, so that the meaning of things is always ephemeral, and multiple ontologies exist in tension, competing, dominating, subverting and altering (Law 2002, Mol 2002).

A focus on phenomenology has proved fruitful. Ingold presents the environment as a ‘generative field’ with which people engage, simultaneously experiencing, dwelling in, assessing and utilising its various elements (2000, 2007a, 2011a, 2011b). Analyses of sensory experience have also underlined the intimacy of human interaction with the material world (Damasio 1999, Howes 2005). The notion of ‘embodiment’ expresses the simultaneously affective and material nature of these interactions and their incorporation at physiological, emotional and imaginative levels (Csordas 1994). Engagement

shifts continuously between subjective phenomenological experience of things and the objective perusal, classification, evaluation and representation of them. Hegel described this as a dialectical process (1979), and Schlanger reminds us of Mauss' point: that in this engagement the body is the first tool, and – critically – the 'technology of the body' includes the cognitive schemes through which people describe and categorise the things with which they interact (2009).

Such categories are not arbitrary: to compose them, people draw directly on the material and behavioural properties of things (Tilley 2007: 17). Analyses of art and material culture have shown how meanings are encoded in objects and practices and communicated inter-generationally, evoking affective responses or aesthetic appreciation, transmitting ideas and practices, inspiring religious fervour and signifying social inclusion (Boivin 2008, Gell 1998, Miller 1998, Morphy 1991, Pinney and Thomas 2001, Tilley 1999). Such approaches can be as readily applied to 'natural resources' as to human-made artefacts (Chen *et al* 2012). Ethnographic research has shown that water is encoded with powerful meanings which draw imaginatively on its particular properties. Thus the literal centrality of water to human life and health, and the recognition that this process is essential to all organic beings, leads consistently to its categorisation as the primary life-generating substance. The co-substantiality between persons and things, and the material flow of water between bodies and environments produces myriad ideas about common substance and connection, for example in imagining the co-identification of people and places (Strang 2002, 2006). And – perhaps most powerfully – the behaviours of water, generated by its particular properties, mean that it is universally employed to articulate notions of flow, transformation, and change over time (Lakoff and Johnson 1980, Strang 2004, 2012).

Implicit in discussions about affect is an acknowledgment of the agency of things. This issue has been much debated. Ingold accepts that things are 'active constituents of a world-in-formation' (2007b: 9, 10), but he rejects the idea of locating agency in them, maintaining that in a 'lifeworld':

> ...bringing things to life… is not a matter of adding to them a sprinkling of agency but of restoring them to the generative fluxes of the world of materials in which they came into being. (2007b: 1)

The notion of flux is appealing, and resonates with Latour's view that agency emerges neither from people nor things but from their combination (1996). Tsing's work on 'friction' (2004) is also useful, describing a generative sea of movement in which people, processes and things collide, abrade, and affect each other.

But these 'collisions and abrasions' do not take place by chance, and the properties of things and persons are neither random nor chaotic: they are the product of (and situated within) long-term,

systemic adaptations and co-development and their properties and effects on each other exhibit some consistent patterns. Tilley argues that the notion of agency is inherent in ideas about affordances and constraints (2007:19, see also Gibson 1979, Gibson and Ingold 1994). In accord with this view, I would suggest that the perceived qualities of things are actively promoted by their particular properties, thus giving things, if not agency *per se,* some agentive force.

Further, I would suggest that the notion of flux, while useful in expressing the dynamic nature of multiple interactions in complex systems, does not give sufficient weight to the materiality of things and the potentially formative (ie. causal) capacities of their particular properties. Nor does it encompass behavioural consistencies in the actions of persons, which may be said to arise, in part, from evolved social behaviours and biological imperatives (which similarly suggest potential causality). This implies a patterned recursive relationship in which both persons and things act upon and affect the emergent properties and behaviours of each other over time. In this sense, they could be said to be co-emergent, mutually producing what Hendry and McLeish describe as 'dependent novelties' in physical and social systems at micro and macroscopic levels (2012).

In this sense a focus on the materiality of things offers some counterbalance to rather absolute post-modern notions of fluidity. If things *and* persons have abiding properties, processes and behaviours which persist over time, and which are formative of their relationships, we could reasonably consider the characteristics of objects (and human and other living kinds), as providing *relative* – though by no means fixed – consistencies in an evolving 'flux' of relations. That is not to say that things, even water, are homogenous: all have their own specific material variations and what Gell described as micro-historical trajectories (1998, see also Appadurai 1986). But there is room for both continuity and change. The perceived qualities of things, and their meanings and uses may change radically according to temporal and cultural context, but their properties also provide consistent experiences and encourage/generate recurrent ideas, values and practices. For example, water's core themes of meaning as a life-generating, life-connecting source, as the basis of wealth, health and power, as a metaphorical base for concepts of movement and flow, recur so reliably in different cultural and historical contexts that there is little choice but to conclude that its material properties are formative in influencing how people engage with water, how they think about it, and how they communicate these ideas and values.

Like any other form of agency, material agency is relational, representing a potentially shifting balance of power. Even objects commonly described as 'inanimate' have particular material properties that compose environments, respond to events, offer potentials for (or challenges to) the creation of artefacts, and provide food for thought. Thus Harman's description of a process of building a tunnel (2009) describes how the engineers involved necessarily had to consider the particular character of the rock, and how it might 'behave'.

Water has readily discernible effects in acting upon people and environments, ranging from immediate molecular interactions in organisms; to carving out watercourses, depositing silt etc. in river valleys; to making vast hydrological movements and transformations at an atmospheric level. In exploring cognitive engagement with these processes, it has been argued that the human imagination is predisposed to formulate concepts epigenetically (Hegel 1979). Thus the persistence of water's properties at every scale enables 'scheme transfers' (Bourdieu 1977) in which ideas about the movement of water through the body are readily transposed to imagine flows of water through local, regional and planetary ecosystems.

At various stages in its movements through persons and cultural landscapes water is conceptualised and treated as a resource. It is, indeed, quintessential in its 'resourcefulness': it has the fluid potential to effect transformations in every sphere of human endeavour, and in this sense constitutes 'uber-potential' in material form. It readily illuminates the relationship between political economy and materiality (see Bakker and Bridge 2006, Hudson 2011). Thus Hudson observes that, while Marx conceptualised economy as containing processes of value creation and value transformation, there is also 'a need to understand production as simultaneously a value creation process *and* a material transformation process' (2011: 2).

I have suggested elsewhere that it can be helpful to frame this process as a shifting continuum of agency (2005). At each stage, things retain their material properties and their agentive capacities, but their re-creation as 'resources' is enabled by the imposition of more and more human (and specifically cultural) agency. As Gell said (1998), this is materially enacted through technology, the 'prosthetic' extension of human physical capacities. Appadurai has highlighted the way that things have a 'social life' as they move through different spatio-temporal contexts (1986). Different cultural environments impose specific meanings and practices on objects, including those framed as 'resources'. But *how* things move is also a fundamentally material business, heavily dependent upon the physical characteristics of the resources in question and on technologies of transport. Ephemeral resources – intellectual property, ideas and so forth – require living people and communicative media. Physically substantial materials such as stone and timber require stalwart oxen, large trucks or hefty ships. Water and other fluids can only be moved in large quantities through channels and pipes. Each material 'resource' thus shapes the technology of its transportation. Also materially relevant is the topography and climate of its location, which may also impede or assist infrastructural developments such as the roads, pipelines, airstrips and canals necessary for their movement.

So in moving through social processes of production, exchange and consumption, the material characteristics of 'resources', the materiality of their original locations and destinations, the material bodies of human beings and – critically – human minds, participate in dynamic and mutually constitutive relationships.

2. Thinking Ethnographically

A (very brief) thumbnail sketch of an ethnographic case study might help to illuminate the shifting balances of agency and causality in this process. Long-term research in south-east Queensland (Strang 2009), provides an account of the Brisbane River, which begins as run-off from the Jimna Ranges, which are today used mostly for cattle grazing and forestry. At the base of the hills, the trickles of water converge into a river that, over millennia, has cut a valley eastwards, swinging across the land in widening loops and from time to time flooding the flat plains it has created. Its movements provided both fertile farmland for colonial settlers, and, with the clearance of land for this purpose, an increasing risk of inundation of the expanding urban areas near the estuary.

Successive Queensland governments have built several large dams to assure drinking water supplies, to irrigate crops and to provide flood mitigation. The shifting balance of human-water agency is clear in that sometimes water subverts this control, disappearing or arriving in unmanageable quantities, and this subversion is combated with deeper bores and desalination plants, flood-prevention infrastructure, pipelines for recycling, and more sophisticated forms of water treatment.

The collection of water for drinking and domestic purposes is the most basic application of human agency. For many millennia prior to the European colonisation of Australia, Aboriginal people cleared and protected springs and dug wells to collect and filter drinking water. Since the first pipeline to the city in 1893, the purpose of the dams and weirs in the Brisbane River Valley has been fundamentally the same: to try to ensure a supply of safe, potable water. Thus water is pumped into treatment plants and filtered; organic biota are killed off with chlorine; fluoride is added, and it is transformed into ‘drinking water supplies’, underlining Miller’s point that even the most basic things that people engage with are rarely ‘raw’ or ‘virgin’ materials, but complex artefacts incorporating sophisticated levels of acculturation (2005).

Via an infrastructural network of pipes, these ‘supplies’ flow down to a familial level and enable the various functions of a domestic context. Here the water flows into an even more immediate scale of engagement at the interface of bodies and water. In an internal cycle of cellular hydration and waste removal: water does things to the body and the body does things to water. Infused with waste, water carries it ‘outwards’ in a variety of ways. Via the larger familial body of the house, it is expelled into sewage pipes and treated. It is then returned to the river for further transformation, in this case via the agency of the environment, in which the river is expected to dilute, absorb or at least carry away the flow of waste.

There is easy coherence between ideas about the ‘right’ balance of water flowing through the body and those about holding and managing it in a wider material environment. Houses, cities, rural

and urban landscapes all depend upon an orderly flow of water and waste. As it runs eastwards, the Brisbane River is regulated with various forms of impoundment. One of the most important of these is the vast 'reservoir' created by the Wivenhoe dam. Irrigators describe such reservoirs as 'liquid gold', highlighting the reality that, for them these are economic 'reserves', and providing a direct demonstration of water's persistent meanings as a source of wealth and generative power. The water from Lake Wivenhoe is diverted through channels and pipes into farm dams, sprayed onto crops, and so transformed into products that express the agency and identity of the region's primary producers. Each product may be said to be composed of 'virtual water': ie. the amount of water required for its production.[1]

Until recently, many of these products were transported down the river to Brisbane. Today this happens primarily by road, but the river still enables the movement of goods outwards. Connecting Brisbane umbilically to larger social and economic bodies, the port provides a conduit for the export of all of the things produced in the river valley. Thus the water that began in the Jimna ranges, now transformed into a range of products, is loaded into containers and carried away across the sea.

The flow of water in south Queensland connects the people, other species and things involved in these activities, and enables ideas about social identity and community along the waterways, and in the State more generally. For example, the Brisbane River is closely involved in the creation and representation of a cohesive identity for the city at its estuary, and local waterways provide the basis for 'neighbourhoods' within its urban sprawl. However, water is always subject to the realities of political economy, and also embedded in these relations are various regimes of inclusion and exclusion. At a local level, these are defined by who has piped supplies; who has licences to abstract, to impound or to harvest water; who has access to it recreationally. Cities and farms consume water that would otherwise feed the ecosystems from which this water is extracted. These social and material relationships extend outwards into larger national and international networks intersecting with a key issue about who owns and controls the flow of potentiality that water represents – and who therefore benefits.

The production of material goods in Queensland entails a further transformation of water. Initially, aquifers, rivers and reservoirs were converted into 'resources' through water allocations. With a prerequisite that they would produce things, landowners were given licences to extract or divert water, first in unlimited quantities and then, as farming intensified, in specifically limited volumetric amounts. In 2000s, however, the Federal Government established a system of water trading that

[1] Anthony Allen's concept of 'virtual water' calculates the amounts of water required to produce crops and other objects. For example, the production of a cup of coffee requires approximately 140 liters of 'virtual water' (Allan 2011, see also Meissner 2012).

converted licences for water allocations into private assets which could be traded on a virtual 'water market'. In this way the ownership and control of water has been commoditised and enclosed by a water-owning elite and, in Polanyi's terms, 'disembedded' from its locality and dispersed into a global economy (1957, see also Howes 1996, Kopytoff 1986). Though virtual, such water has not entirely evaporated: it has simply held elsewhere – potentially anywhere. But most will flow in virtual form into the asset reserves of transnational corporations and, via these into the streams of wealth circulating the globe and connecting local, national and international economies. Where it pools and who benefits from it depends on the vagaries of the market and a shifting balance of international relations.

What transpires when water 'resources' flow out of local social and ecological systems into levels of consumption – in cities, in other countries – that the originating material environments cannot sustain? In Queensland, the rapid intensification of farming and dependence on irrigation has led to chronic over-use of water, creating a host of detrimental effects on social and ecological systems both locally and further afield. Thus the impacts of its human population, like Queensland's water, are carried outwards into larger scale human-environmental interactions.

At each level, human actions upon things have wrought changes in their properties and behaviours: for example, soil degradation and pollution along the Brisbane River have compromised the quality of water, impinging upon its chemical composition and disabling its capacities to support aquatic biota and other participants in ecological processes. Simultaneously, anxieties about pollution have shifted people's perceptions of water as a life and health giving substance, and initiated changes in behaviour, for example leading to the purchase of bottled water.[2] As well as degrading the immediate material environment, the State's increasing human population is also burning fuel, consuming other resources (often imported from other environments), and contributing to the production of CO_2.

3. Emergent Relations

The effects of such patterns of resource use and consumption include major social and political contests for resources, and global environmental change: for example, the melting of arctic ice caps and of the glaciers on which whole societies and economies depend. Such events lead to questions about sustainability and how particular modes of human-environmental engagement can destabilise long-term human-environmental relationships to the point where shifts – novelty – in material

[2] The production of each bottle uses approximately six times the amount of water that it contains, as well as energy (fuel) and materials that also have to be abstracted from particular locations and transported, with commensurate ecological effects.

processes can occur. Anthropological and ecological work on ‘resilience’ (Hastrup 2011, Humphrey and Sneath 1999) underlines the potential flexibility in such engagements, but this is also an acknowledgement of shared human and non-human agency, and of the reality that this takes material form.

Archaeological, historical and ethnographic accounts suggest that the systemic order provided by consistency in the properties and behaviours of persons and things, and by an evolving balance of agency/causality, is tenuous. Sufficient perturbation can produce tipping points that initiate unmanageably rapid changes in the constellation of the multiple relationships between social and ecological systems (Bentley and Hahn 2011, Diamond 2005, Gladwell, 2000). A key question is the extent to which the abiding properties of things and persons can reassert sustainable relations: ie. those in which the participants do not become casualties. For example, accelerated rates of species extinction, according to the International Union for Conservation of Nature (IUCN), are now occurring at a rate of at least 1000 times (and possibly 10,000 times) that which pertained prior to the last two centuries.

This highlights the issue of pace and the adaptive capacities of the participants. Recent literature on social and environmental change highlights not only emergent practices but an overarching change in pace. Virilio frames this as acceleration (see Armitage 2001); Harvey describes space-time ‘compression’ (1990), and Hylland-Erikson explores the notion of ‘overheating’, which resonates symbolically with representations of global warming (pers. comm.). There are useful questions about how such changes are underpinned by growth-oriented ideologies, and the extent to which these express long-term behavioural imperatives which may have had long-term evolutionary advantages, but which – with finite material resources – may now be maladaptive. There are multiple (often subjective) criteria for describing systemic ‘failure’ or ‘development’. The notion of emergence seems objective at a material level, but applied to human-environmental relationships, and even in more abstract terms to ‘mind’, it rapidly intersects with questions about consciousness, moral choice, ideology and values.

These questions are beyond the remit of this paper, but in informing notions of emergence, they do lead to consideration of whether the properties of things and persons function to stabilise their interrelationships by keeping a brake on radical change, thus preventing emergence from becoming emergency. Alternatively, or perhaps simultaneously, such properties may also be said to limit the adaptive capacities of participants

4. Conclusion

The ethnographic mapping of events highlights the point that ‘causality’ or ‘agency’ in

human-environmental relations is dynamic and negotiable. In complex systemic interactions, the balance of agency shifts rapidly and continuously between people and things/material environments over time. Within such interactions, there are multiple micro-relations in which causality and its subveniences and dependencies may be defined. At every level of engagement, underpinned by ‘laws’, the material and behavioural properties of objects and persons provide some consistency, but they also act upon and transform each other in a mutually constitutive and dynamic process of co-evolution, or as Deleuze would put it ‘becoming’ (1994).

These terms imply a potential for novelty to ‘emerge’ from the complex interactions between people and material things. It also implies a non-anthropocentric measure of equality, and in thinking about causal dependencies in human-environmental relations, it may be useful to subvert conventionally dualistic notions of nature and culture. In science, things and persons are seen as distinct. A focus on physical causality frames 'effects' as ‘the manifestations of the characteristic causal powers or dispositions of objects’ (Hendry and McLeish 2012). Metaphysical debates have tended to focus on rigorous philosophical abstractions about ‘mind’, suggesting an implicit separation from the physicality of being.

But the assumption of an intrinsic divide between persons and things can act as a barrier to articulating the recursive flows between material and mental domains. Anthropological theory, much influenced by non-Western worldviews, has moved steadily towards more integrated concepts in which phenomenological and sensory experiences of the world are materially and imaginatively ‘embodied’, and personhood is conceptually and practically extended to objects and environments (Descola and Palsson 1996). In this model ‘culture’ and ‘nature’ permeate each other, dissolving the conceptual boundaries between things and persons.

Approached from this theoretical perspective, the ethnographic mapping of human-environmental engagements seems to offer a bridge between emergent material and mental processes. It can provide – albeit in very broad terms – ways of thinking about an experiential ‘causal history’ that is formative of the ways that ‘systems of neurons support a mental life’ (Hendry and McLeish 2012). Hopefully, this opens up some potential pathways towards understanding the co-emergence of mind and matter.

References

Allan, J.A. 2011. *Virtual Water: tacking the threat to our planet’s most precious resource,* London: I. B. Tauris.

Appadurai, A. 1986. *The Social Life of Things: commodities in cultural perspective,* Cambridge: Cambridge University Press.

Armitage, J. (ed) 2001. *Virilio Live.* London: Sage.

Bakker, K. and Bridge, G. 2006. 'Material Worlds? Resource Geographies and the "Matter of Nature"', in *Progress in Human Geography*, 30 pp 5-27.

Bennett, J. 2009. *Vibrant Matter: a political ecology of things,* Durham, NC: Duke University Press.

Bentley, R.A. and Hahn , M.W. (2011) 'Is There a 'Neutral Theory of Anthropology'? (Comment on Lansing and Cox)', in *Current Anthropology.*, 52 (1). pp 117-118.

Boivin, N. 2008. *Material Cultures, Material Minds: the impact of things on human thought, society, and evolution,* Cambridge, New York: Cambridge University Press.

Bourdieu, P. 1977. *Outline of a Theory of Practice,* Transl. R. Nice. Cambridge: Cambridge University Press.

Busse, M. and Strang, V. 2010. 'Introduction', in *Ownership and Appropriation,* ASA Monographs. Oxford, New York: Berg. pp 1-20.

Chen, C. Macleod, J. and Neimanis, A. (eds) 2012. *Thinking With Water,* Montreal: McGill-Queens University Press.

Deleuze, G. and Parnet, C. 2006. *Dialogues II.* Transl. Hugh Tomlinson, Barbara Habberjam, London: Continuum.

Deleuze, G. 1994. *Difference and Repetition,* Transl. Paul Patton. London: Athlone.

Damasio, A. 1999. *The Feeling of What Happens: body and emotion in the making of consciousness*. New York: Harcourt Brace.

Descola, P. and Palsson, G. (eds) 1996. *Nature and Society: anthropological perspectives,* London, New York: Routledge.

Diamond, J. 2005. *Collapse: How Societies Choose to Fail or Succeed,* New York: Viking.

Gell, A. 1998. *Art and Agency: an anthropological theory,* Oxford: Clarendon Press.

Gibson, J. 1979. *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Gibson, R. and Ingold, T. (eds) (1994) *Tools, Language and Cognition in Human Evolution.* Cambridge: Cambridge University Press.

Gladwell, M. 2000. *The Tipping Point: How Little Things Can Make a Big Difference* London: Abacus.

Haraway, D. 2008. *When Species Meet,* Minnesota, MN: University of Minnesota Press.

Harman, G. 2009. *Prince of Networks: Bruno Latour and metaphysics.* Melbourne: Repress.

Harvey, P. 2012. 'Knowledge and Experimental Practice: a dialogue between Anthropology and Science and Technology Studies', in R. Fardon, O. Harris, T. Marchand, M. Nuttall, C. Shore, V. Strang and R. Wilson. (eds) *A Handbook of Social Anthropology,* Association of Social Anthropologists of the United Kingdom and Commonwealth, London: Sage. pp.?? in press.

Harvey, D. 1990. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Cambridge, MA: Blackwell.

Harvey, P. and Knox, H. 2010. 'Extraction, Materiality and the "Science of the Concrete" in Engineering Practice', in T. Bennett and P. Joyce (eds), *Material Powers: cultural studies, history and the material turn,* London: Routledge, pp. 124−141.

Hastrup, K. 2011. *Waterworlds: natural environmental disasters and social resilience in anthropological perspective,* Mid-term activity report, University of Copenhagen: European Research Council.

Hegel, G. 1979. *The Phenomenology of Spirit,* translated by A. V. Miller Oxford: Oxford University Press.

Henare, A., Holbraad M. and Wastell, S. 2007. *Thinking through Things: theorising artefacts ethnographically,* London: Routledge.

Hendry, R. and McLeish, T. 2012. *Durham Emergence Project*, Draft Proposal.

Howes, D. (ed) 1996. *Cross-cultural Consumption: global markets, local realities,* London, New York: Routledge

Howes, D. (ed) 2005. *Empire of the Senses: a sensual culture reader*. Oxford, New York: Berg.

Hudson, R. 2011. 'Critical Political Economy and Material Transformation' in *New Political Economy,* London: Routledge. e-publication DOI:10.1080/13563467.2011.578736

Humphrey, C. and Sneath, D. 1999. *The End of Nomadism? Society, State and the Environment in Inner Asia,* Cambridge: White Horse Press.

Ingold, T. 2000. *The Perception of the Environment,* London, New York: Routledge.
—2007a. 'Earth, Sky, Wind, and Weather' in C. Low and E. Hsu (eds), 'Wind, Life, Health: anthropological and historical perspectives, in *Journal of the Royal Anthropological Institute,* Special Issue. pp 19-38.
—2007b. 'Materials Against Materiality', discussion article in *Archaeological Dialogues,* 14(1) pp 1-16.
—2011a. *Being Alive: essays on movement, knowledge and description,* London: Routledge.
—2011b. 'Walking with Dragons: an anthropological excursion on the wild side' Firth Lecture, ASA Conference, University of Wales, Trinity-St David. September 15th 2011.

International Union for the Conservation of Nature (IUCN)
http://www.google.co.uk/search?hl=en&q=iucn+species+extinction&meta=

Knappett, C. 2007. 'Materials With Materiality?', In *Archaeological Dialogues,* 14(1) pp 20-23.

Kopytoff, I. 1986. 'The Cultural Biography of Things: commoditization as process', in A. Appadurai (ed), *The Social Life of Things: commodities in cultural perspective,* Cambridge pp 64-91.

Lakoff, G. and Johnson M. 1980. *Metaphors We Live By,* Chicago, London: University of Chicago Press.

Latour, B. 1996. 'Lettre à mon ami Pierre sur l'anthropologie symétrique' Culture Materiélle et Modernité, in *Ethnologie Française.* 26 (1) pp 32-37.
— 2005. *Reassembling the Social: an introduction to Actor Network Theory*, Oxford: Oxford University Press.

Law J, 2002. *Aircraft Stories: decentring the object in technoscience,* Durham USA: Duke University Press.

Margulis, L. and McMenamin, M. (eds). 1992. *Concepts of Symbiogenesis: historical and critical study of the research of Russian botanists,* New Haven: Yale University Press.

McMenamin D. and McMenamin M. 1994. *Hypersea,* New York: Colombia University Press.

Meissner, S. 2012. 'Virtual Water and Water Footprints: global supply and production chains and their impacts on freshwater resources' in P. Hahn, K. Cless and J. Soentgen (eds). *People At The Well: kinds, usages and meanings of water in a global perspective,* Frankfurt, New York: Campus Verlag. pp 44-64.

Miller, D. 1998. (ed.) *Material Cultures: why some things matter.* Chicago: University of Chigago Press.
— (ed) 2005. *Materiality: politics, history and culture,* Durham NC.: Duke University Press.

Mol, A. 2002. *The Body Multiple: ontology in medical practice,* Durham USA: Duke University Press.

Morphy, H. 1991. *Ancestral Connections: art and an Aboriginal system of knowledge,* Chicago: Chicago University Press.

Pinney, C. and Thomas, N. 2001. *Beyond Aesthetics: art and the technologies of enchantment,* Oxford: Berg.

Polanyi, Karl. 1957. *The Great Transformation,* Boston: Beacon Press.

Ritvo, H. 1987. *The Animal Estate: the English and other creatures in the Victorian Age,* Cambridge Mass.: Harvard University Press.

Schlanger, N. (ed) 2009. *Techniques, Technology and Civilization/ Marcel Mauss,* New York and Oxford: Durkheim Press, Berghahn Books.

Strang, V. 2002. 'Life Down Under: water and identity in an Aboriginal cultural landscape', *Goldsmiths College Anthropology Research Papers.* No. 7. London: Goldsmiths College.

― 2004. *The Meaning of Water,* Oxford, New York: Berg.

― 2005. 'Knowing Me, Knowing You: Aboriginal and Euro-Australian concepts of nature as self and other', in *Worldviews* 9.1. pp 25-56.

― 2009. *Gardening the World: agency, identity, and the ownership of water.* Oxford, New York: Berghahn Publishers.

― 2012. 'Conceptual Relations: water, ideologies and theoretical subversions', in C. Chen, J. Macleod and A. Neimanis (eds) *Thinking With Water,* Montreal: McGill-Queens University Press.

Strang, V. and Garner, A. (eds) 2006. *Fluidscapes: water, identity and the senses,* Special issue of *Worldviews*. Vol 10. No 2. Brill Academic Publishers.

Strathern, M. 1996. 'Cutting the Network' in *Journal of the Royal Anthropological Institute,* 2(3) pp 517-35.

Tilley, C. (ed.) 1999. *Metaphor and Material Culture,* Oxford: Blackwells.

— 2007. 'Materiality in Materials', in *Archaeological Dialogues*, 14 (1) pp 16-20.

Tsing, A. 2004. *Friction: an ethnography of global connections,* Princeton: Princeton University Press.

Verdery, K. and Humphrey, C. 2004. *Property in Question: Value Transformation in the Global Economy*, Oxford: Berg

Vernadsky, V. 1986. *The Biosphere,* Santa Fe NM: Synergetic Press.

Contact Details: Veronica Strang
Institute of Advanced Study, Durham University
E-mail: veronica.strang@durham.ac.uk

22

Agent-based simulation of emergence in populations of financial organizations

Dr Philip Garnett[1] & Dr Simon Mollan[2]
Durham University[1]
University of Liverpool[2]

Summary: As the vast majority of businesses leave no trace and therefore history is often written from the point of view of the winners. This is especially true of the British banking sector, despite the existence of rich archives. History therefore tends to explain the growth and success of British banks in terms of the banks' own operations. We challenge this, and with the aid of a unique data set and agent-based modeling techniques propose an alternative view that the banking sector that we see today emerged from a process of merger an acquisition. Further more we suggest individual banks had limited agency in terms of their strategy, and that population level forces shaped the banking sector that we see today.

1. Introduction: History of large companies (such as banks) often discusses the success or failure of an organization in terms of the strategy and decisions taken by its directors. There is also an archival bias towards successful companies. This can lead to post-hoc justification for the success of an organization. Where significance is attached (in hindsight) to decisions, or perhaps even individuals, to explain the success of the company in question (1,2). The purpose of this paper is to use the population dynamics of the British banking sector to challenge this view by providing and alternative hypothesis for the development of populations of large organizations, a hypothesis where the environment is key.

We have developed a unique dataset that tracks the development of the majority of British banks from the 17th century to the present day. Taken largely from the Register of Bank Name Changes & Liquidations (3) not only is the dataset unique in its size and scope, it extends to over 3000 banks, we also linked the activities of the individual banks together. This makes it possible to track the development of any individual bank through time. Originally stored in a relational database, we imported the database into a graph database (4). Graph databases store data in the form of nodes and edges linking nodes. In the case of our dataset a node is a bank, and the attributes of the node will include details such as (including but not limited to) its name, creation date, end date, and current status. The nodes in the graph database are linked together by edges that store relationships between nodes. The edges linking our bank nodes indicate what the relationship is between two banks. For example, the relationship between two banks could be 'taken over', indicating that one back consumed the other. Storing the data in this manor allows us to produce 'family trees' for all banks in the data set. Figure 1 shows a section of a 'family tree' for the National Provincial banking group.

From the dataset we can discover various details about the population dynamics of the British Banking sector. Figure 2 shows the changing population of banks through time. It clearly shows that between 1600 and 1810 the population of British banks was increasing exponentially with a growth rate of about 2.7%. After 1810 something significant changes and for the next 200years the population of British banks exponentially decreases at a rate of about 1.2% until we reach the situation that we are in today. In the UK today there are approximately 100 active banks, and 5 very large banks dominate the

sector. The period of time after 1810 is the focus of this paper.

The shape of the plot in figure 2 is reminiscent of an ecological population collapse. Population collapses can be triggered by a number of events. One possibility is that a population expands far beyond its carrying capacity causing food resources run out and the population ultimately crashes. Equally the impact of a changing environment could cause a population to crash (5). It is difficult to make a simple comparison with banks and biology. During this time the UK GDP was (and in general still is) increasing (6), which might suggest the crash in the number of banks was some other environmental factor, and not simply resource. It still suggests that some sort of system level behavior change drove the population of British banks.

Shortly after 1810, in 1825, banking regulation changed. Prior to this date it was difficult for banks to expand in size, as they were limited to 5 partners. In 1825 the rules were changed and banks were able to expand through merger and acquisition (7). We therefore hypothesize that this regulatory change set the banking sector on a trajectory that brought it to where it is today, what we see today is an emergent property of a population of individual banks that are all expanding through merger and acquisition as fast as they can. Once a small number of banks started to grow using this method it essentially forced other banks to follow suit. In order to remain competitive in the sector all banks were compelled to follow this pattern of behavior, follow the herd. Failure to do so would risk being left behind, or being consumed by a larger bank. If correct, this hypothesis would suggest that the banks that came out of this process did so more out of luck, or indeed random chance, and different names could quite easily be posted above their doors.

Emergence can be used to describe a phenomenon where an unforeseen system level behavior is produced by the interactions of individual agents within that system (8). In order to test whether our hypothesis that the development of the British banking sector post 1810 what dominated by merger and acquisition behavior of individual banks. Resulting in the sector being dominated by a small number of very large banks that is often described as a cartel, we developed an agent-based model of a population of banks. In the model banks are able to carry out three simple behaviors at each time point in the model. At some probability a new bank can be created, an existing bank can fail, or a bank can attempt to merge with another bank selected at random from the population. The time-step of the simulation is one month. The models were developed using the CoSMoS process (9).

Using rates from the real data this simple setup is sufficient to reproduce the general population trends seen in the real data. Figure 3 shows the simulated population of banks plotted over the top of the real data. It is worth noting that we are not trying to exactly reproduce the real data, but more discover something about the underlying properties of the system.

2. Discussion: We are unable to discount our hypothesis that the change in the regulatory environment around 1825 dominated the population of banks for the next 200 years. Using rates from the real data for the creation, failure and merger of banks we are able to reproduce the general population trends seen in the population of British banks. We suggest that the population (or system) level behavior of an exponentially decreasing population is an emergent property banks being able to grow through merger and acquisition. The behavior was deemed to be successful and therefore other banks in the

population copied the behavior. This can be described as mimetic behavior, were one organization imitates the ‘successful’ behavior of another. It may also be an example of institutional isomorphism, where the units in a population come to resemble each other as a result of environmental factors.

The exact timing of the events is interesting. The regulatory change that gave the banks more freedom to expand through merger and acquisition did not come into force until 1825 (7). By this time the population of banks had already been decreasing for approximately 15years. One possible explanation is that banks were already finding ways to expand, and that the regulatory change might have been in response to pressure from banks. There might also be evidence that the banking environment could have forced banks to look for ways of increasing their resilience. This is now an active area of study.

There is an interesting thought experiment that can be carried out. Complex systems techniques of data analysis, modelling and simulation potentially have a lot to offer when applied to the testing of historical hypothesis. The problem comes with validating the results from the models and simulations. Our data and simulation is so far unable to refute our hypothesis that does not mean it is necessarily correct. There is the possibility for more comparisons to be made between the simulation results and the real data that might provide additional insight into the validity of basic assumptions and the hypothesis itself. If we do become satisfied that our models are valid would anyone trust their predictive capacity? If in 1870 this model was presented and the prediction was made that the British Banking sector was heading towards a small population of banks, dominated by a cartel of very large banks. Would anyone have believed it?

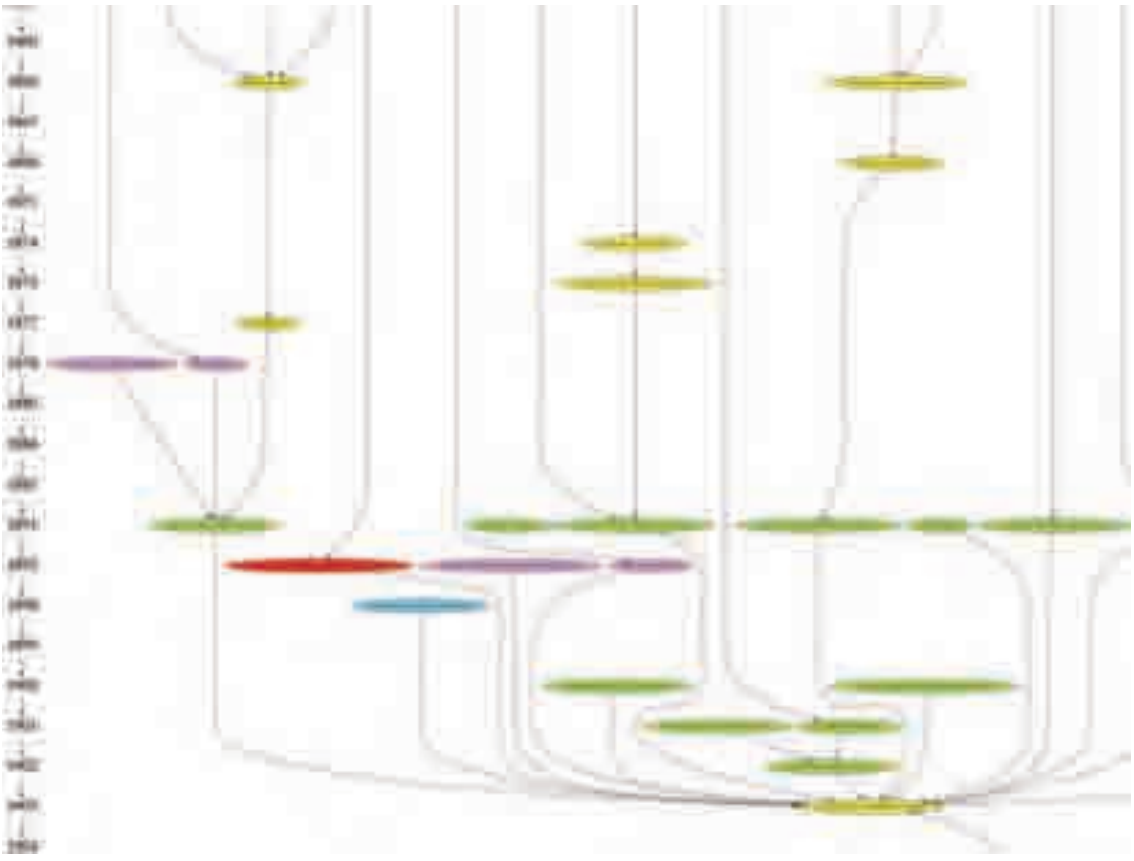


Figure 1: Section of the National Provincial Family Tree.

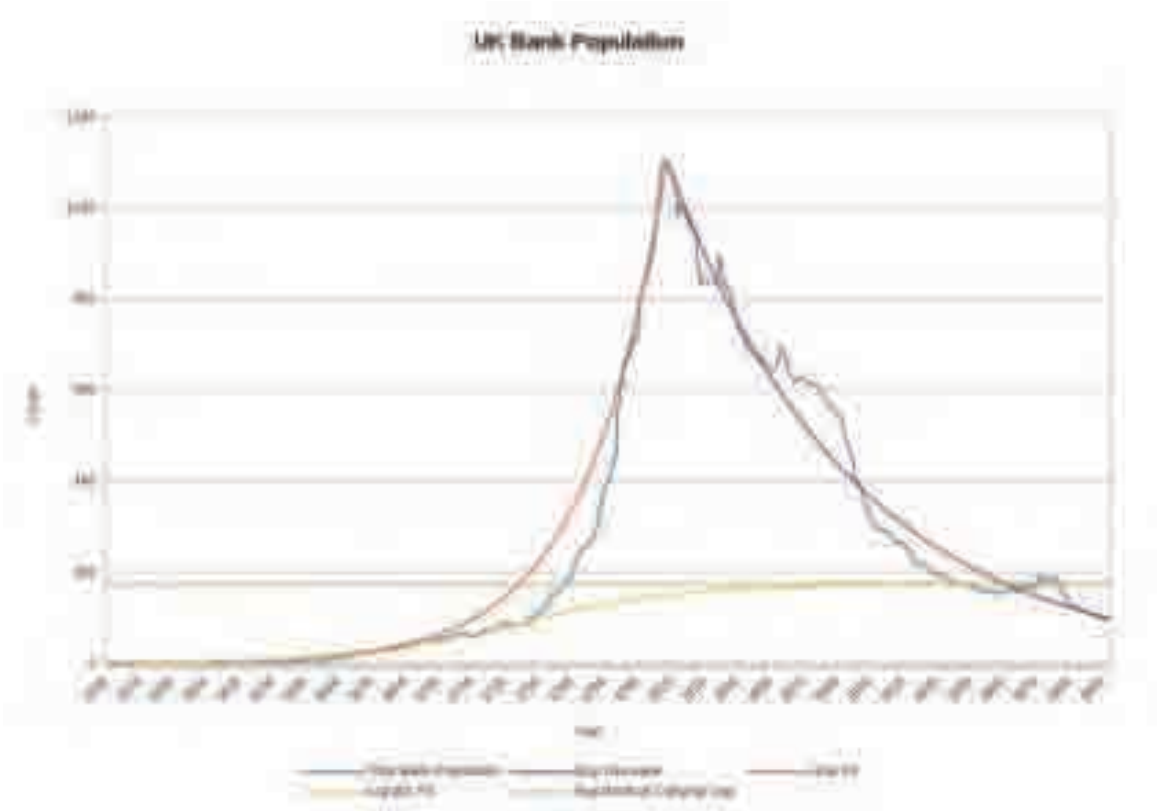


Figure 2: Changing population of UK banks. The population increases up to a peak of 1100 banks in 1810 before declining for the next 200 years.

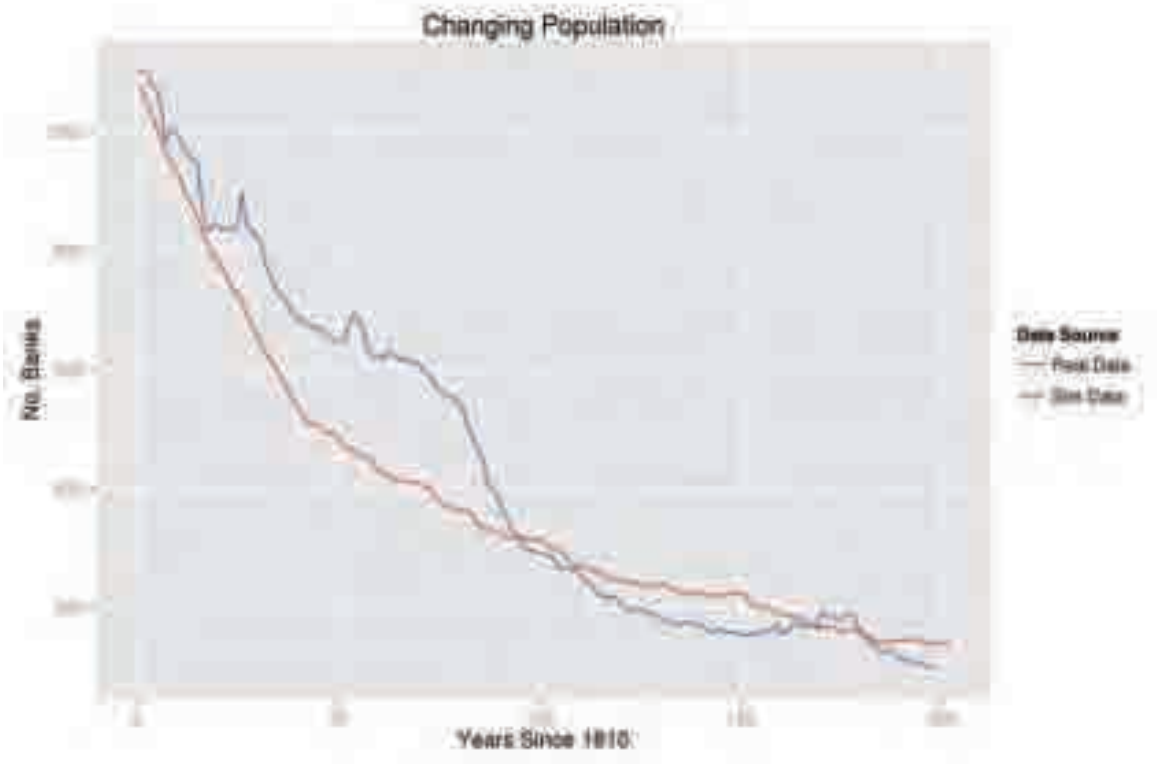


Figure 3: The real and simulated banking population from 1810 to the present day. The simple simulation

reproduces what we see in the real data with the minimum of assumptions.

Acknowledgements: The funds that made this work possible were provided by the Leverhulme Trust as part of the Tipping Points Project at Durham University. We would like to thank our colleagues on the Tipping Points Project.

Reference:

1. Ackrill, M. and L. Hannah (2001). Barclays: The Business of Banking, 1696-1996. Cambridge, Cambridge University Press.
2. Homes, A. R. and E. Green (1986). Midland. 150 Years of Banking History London, Batsford.
3. Moore, C. (1992). Register of Bank Name Changes and Liquidations. East Grinstead, Reed Information Services Ltd.
4. Neo4J (2012). "The World's Leading Graph Database." Retrieved 17-10-2012, 2012, from www.neo4j.org.
5. Elton, C. S. (1924). "Periodic Fluctuations in the Numbers of Animals: Their Causes and Effects." The Journal of Experimental Biology **2**: 119-163.
6. Solomou, S. and M. Weale (1991). "Balanced estimates of UK GDP 1870-1913." Explorations in Economic History **28**(1): 54-63.
7. Sykes, J. (1926). The Amalgamation Movement in English Banking, 1825-1924. London, P.S. King and Son Ltd.
8. O'Connor, T. (1994). "Emergent Properties." American Philosophical Quarterly **31**(2): 91-104.
9. Andrews, P. S., F. Polack, et al. (2010). The CoSMoS Process version 0.1., University of York.

23 Qatar and Japan: A Marriage of Convenience?

David Roberts
Durham University

Summary: Centered on Qatari oil exports to Japan, Qatari-Japanese relations go back decades. Subsequently, with Japan looking to diversify its fuel importing mix and the discovery of the world's largest gas field mostly in Qatari territory, after regional and financial crises intervened, Japan played the central role in funding the establishment of Qatar's Liquefied Natural Gas (LNG) industry. Today, Qatar's LNG industry has transformed the country and made Qatar per capita the richest country on earth. Two decades on and Japan still retains a particular affection from many in the Qatari elite.

However, there are challenges ahead for this crucial bi-lateral relationship. The gas market is changing significantly with old exporters reserving their gas for domestic markets (Indonesia) while other exporters come to the fore (Australia). America too is reveling in the shale gas boom, which has the potential to turn the whole market upside-down. On the demand side there is, as ever, huge competition in the Asia Pacific region. China's plans are to significantly increase its imports of gas, which could cause a significant supply constriction if some of the US and Australian plans do not come on line; a perfectly reasonable fear. As for Japan, the ramifications of the Great East Japan Earthquake are yet to be fully understood. However, there was short-term peak in LNG demand which is expected to continue to the longer term. Lastly, there is a changing of the guard in Qatar. The elite that undertook the deals with Japan are slowly retiring and with it – some may fear – is Qatar's unusual affection for Japan.

Yet quite clearly Japan for one are actively bolstering their relationship with Qatar. A combination of existing personal relations between the two countries being passed on, Japan's solid, trusted reputation as a good, financially-sound buyer and Qatar's slick LNG operation and prodigious growth potential highlight that this bilateral relationship is well grounded to counter the challenges ahead.

Introduction

2012 has officially been dubbed the year of Qatari-Japanese relations marking the 40th year anniversary of the establishment of official diplomatic relations. In Doha there have been tens of Japanese cultural events linked to Japanese cinema, Japanese dance, and Japanese art. Similarly in around Japan there have been events put on exposing Japan to elements of Qatari culture. Indeed, no country has done more cultural outreach in Qatar than Japan in recent years. With imitation being the highest form of flattery, this idea of a 'year' of bilateral celebrations has been deemed to be so successful that the UK are aiming to make 2013 the year for Qatari-UK relations.

Qatar-Japanese relations go back until to 1972 if not before unofficially. Japan has imported Qatari oil and engaged in wider trade for decades and more recently the importance of Qatar to Japan as a supplier of Liquefied Natural Gas (LNG) has increased dramatically. Moreover, it was the crucial supportive and financial role that Japan played in the establishment of the LNG industry in Qatar that has cemented bilateral relations. Subsequently, Qatar has become a key supplier of LNG to Japan; an essential part of Japan's domestic energy mix.

However, despite these strong foundations, there are a series of challenges ahead. The longer term ramifications of the East Japan disaster are yet to be fully understood in the context of Japan's energy make-up. New supply challengers for Qatar stemming from Australia in supplying gas and from evolving shale gas technologies are also yet to be fully worked out. Despite these concerns, the fundamental traits that solidified the relationship in the beginning – strong elite relations; Japan's profile as a trusted, financially stable buyer; and Qatar's economies of scale – are likely stand this key bilateral relationship in good stead in the coming decades.

The Energy Link

Japan's level of energy self-sufficiency is notably low at only 16% and without nuclear power it would be at 4%.[1] This has left Japan dependent on energy imports for decades. The volatility in the oil market in the 1970s and 1980s prompted Japan to diversify its energy mix to a greater degree. Subsequent policies reduced Japan's dependence on oil from 98% in 1970 to 61% in 1999.[2]

---

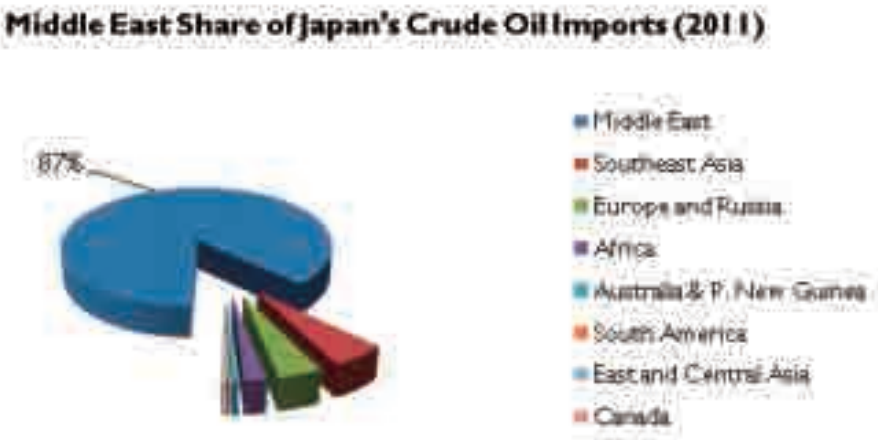
[1] Rami Abdulkarim, "Natural Gas Security of Supply in Japan," *Shingetsu Electronic Journal of Japanese-Islamic Relations* 5 (March 2009): p.16.

"Japan," in *Country Analysis Briefs* (Washington: Energy Information Administration, 4 June 2012), p.1.

[2] All figured cited in Hayden S Lesbirel, "Diversification and Energy Security Risks: The Japanese Case," *Japanese Journal of Political Science* 5, no. 1 (May 2004): p.4.

However, though the absolute dependency on oil was reduced, the relative dependency on Middle East oil has increased to 87% in 2011.[3]

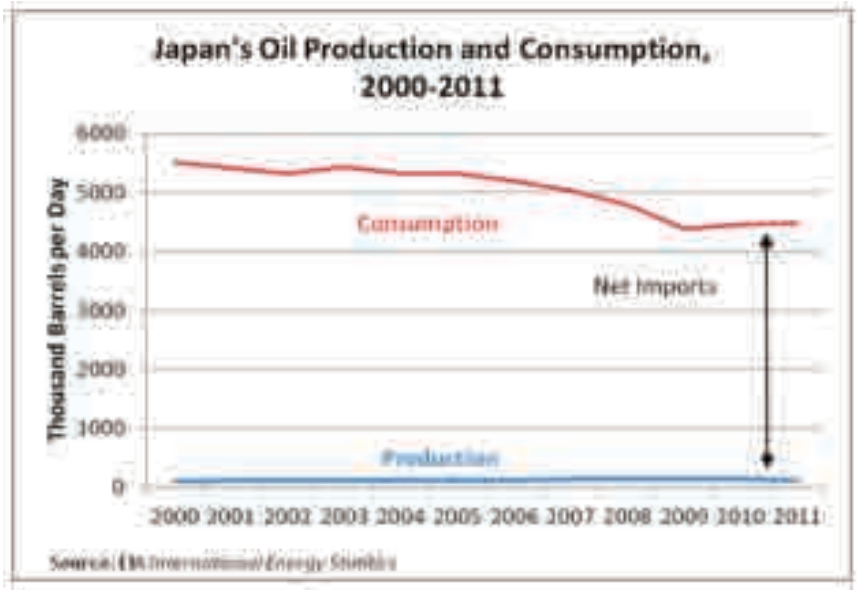
Figure 1 : The Middle East Share of Japan's Crude Oil Imports



Source: Japan's New Energy Frontier and the Middle East[4]

Japan's domestic oil reserves are negligible while per day Japan consumes the third largest amount of oil in the world; 4.5 millions of barrels per day. This leaves a severe gap that needs to be filled by net imports.

Figure 2 : Japan's Oil Production and Consumption



Source: Graph taken from EIA Country Analysis[5]

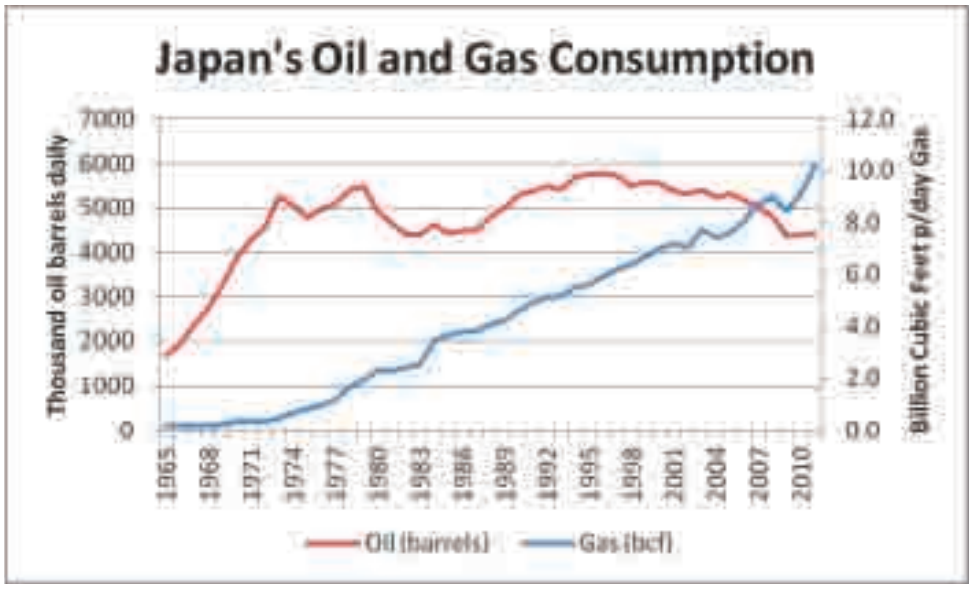
---

[3] John Calabrese, "Japan's New Energy Future and the Middle East," *The Middle East Institute* (11 June 2012), http://www.mei.edu/content/japan%E2%80%99s-new-energy-future-and-middle-east

[4] Ibid.

Since the 1990s in particular, gas in the form of LNG has become increasingly important in Japan's energy mix. However, as with oil, Japan's domestic production is negligible and in the past 40 years Japan has yet to produce more than 5% of the gas it consumes domestically.[6] Hence, the rise in Japan's gas imports over the past three decades to lessen the dependence on oil.

Figure 3 : Japan's Oil and Gas Consumption



Source: BP Statistical Review of World Energy[7]

From the inception of the industry in the 1960s, Japan understood the potentially useful role that LNG could play. In the early years, however, the price was not that appealing as the industry was in its expensive infancy. Yet as the industry progressed, Japan invested ever more and secured contracts. Today, Japan has a world-class LNG infrastructure with at least 27 regasification terminals amounting to almost half of the world's regasification capacity.[8] Historically, Brunei, Indonesia, and Malaysia in particular have been the key suppliers.

---

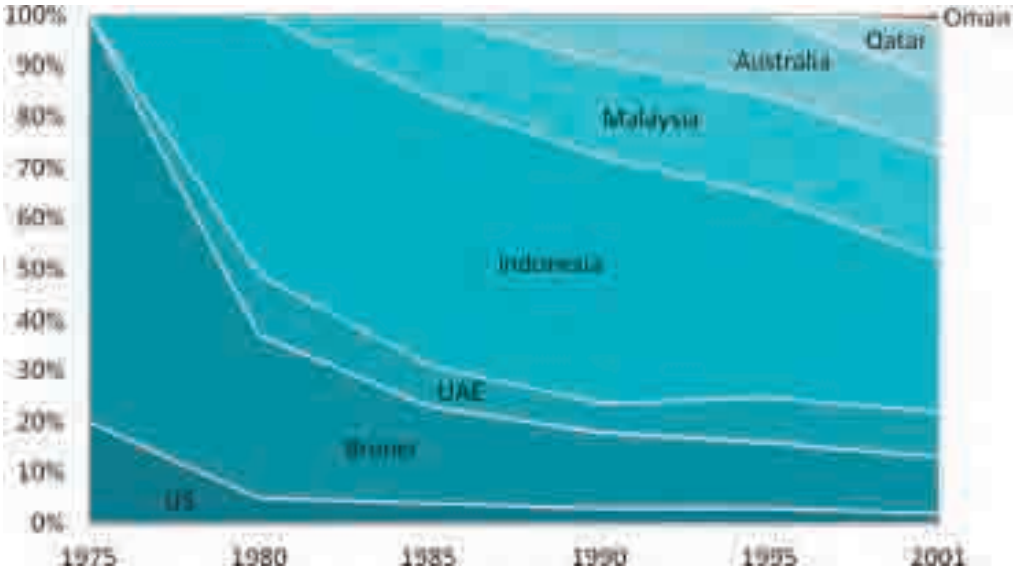
[5] "Japan," p.3.

[6] Sam Stone, "Gas & Geopolitics: The Foreign Policy Implications of Energy Import Dependency" (Stanford University, May 2010), p.52.

[7] *Bp Statistical Review of World Energy*, (London : BP, June 2012).

[8] Abdulkarim, "Natural Gas Security of Supply in Japan," p.39.

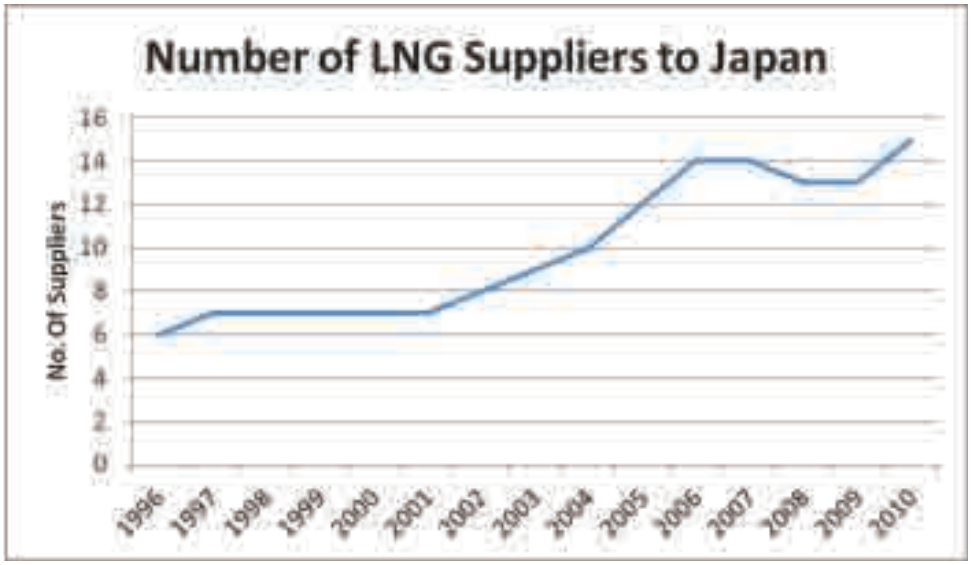
Figure 4 : Japanese Sources of LNG 1975-2001



Source: Natural Gas Security of Supply in Japan[9]

However, Japan has diversified the number of countries on whom it relies for imports in recent years.

Figure 5 : Number of LNG Suppliers to Japan

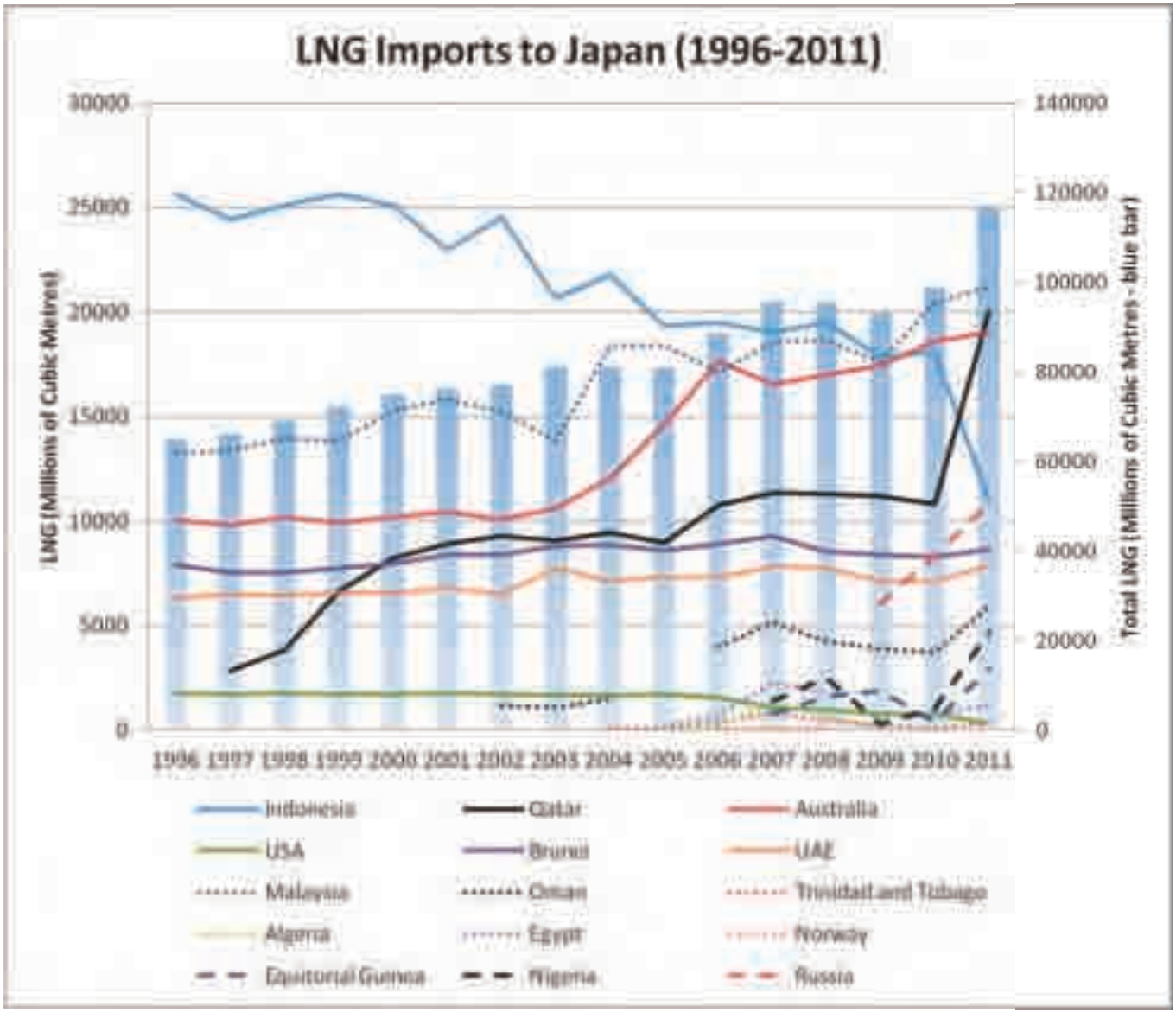


Source: BP Statistical Review of World Energy[10]

[9] Ibid.: p.40.

[10] *Bp Statistical Review of World Energy*.

Figure 6 : LNG Imports to Japan



Source: BP Statistical Review of World Energy[11]

Of particular note in this graph is that the rise in Japan's LNG imports (blue bar) has been accompanied by a severe decline in Indonesian LNG (blue line) and a precipitous rise in Qatari LNG from 2010. Equally, Malaysia (dotted dark blue line) continues to play a key role as supplier, while Australia (red line) has been steadily increasing supply.

Lastly, it should be noted that there is no central LNG buyer in Japan. The energy market is highly fragmented with tens of different companies purchasing LNG independently. The fragmentation is due largely to the difficulties of linking different areas of Japan together given the mountainous terrain.

[11] Ibid.

Qatar

Qatar has been exporting oil since the 1950s but is relatively new to the world of LNG. The world's largest gas field was discovered off the north coast of Qatar in 1971. Most of the field is under Qatar's jurisdiction, but a significant portion belongs to Iran; a potentially troublesome complication.

Figure 7 : Map of Qatar and the North Field/South Pars Gas Field



Source: EIA Qatar Country Background[12]

The field was not developed until the early 1990s due to political instability in the region and domestic political and financial difficulties. Abdullah Al Attiyah was the key interlocutor for Qatari-Japanese relations. He was and remains today one of the Emir's most trusted allies and was made Energy Minister in 1992.

In the 1980s with Japan's economy in the ascendency, increasing energy demand, and with significant experience in developing LNG sites, various Japanese consortiums showed an interest in Qatar's potentially enormous reserves. Also, with South Korea and Taiwan entering the LNG market in 1986

[12] "Qatar: Background," in *Country Background* (Washington DC: U.S. Energy Information Administration, January 2011).

and 1990 respectively, there was greater impetus for Japan to tie up longer term contracts.[13]

Central to Japan's success in Qatar (and elsewhere) was the influence and skills of various *Sogo Shosha*, what may be loosely described as general trading companies. *Sogo Shosha* such as Mitsui Bussan, Marubeni, and the Mitsubishi Corporation acted as "the glue" between the State and its gas companies and the Japanese regional buyers.[14]

Despite a crisis ensuing in 1992 when BP exited the Qatari market, Mobil stepped in soon after and agreements were signed. The key initial end buyer was Chubu Electric which signed up in May 1992. In addition to Japanese end customers, Japanese *Sogo Shosha* brokering the deals, Japanese firms involved in the construction of the apparatus, the Japanese Government oversaw key financing. The Government-led Export-Import Bank of Japan (J-EXIM) provided the majority of the $3bn for the first LNG 'train' in Qatar along with other Japanese banks. Eventually, in January 1997, twenty six years after the discovery of the North Field, the first LNG tanker arrived in Japan.

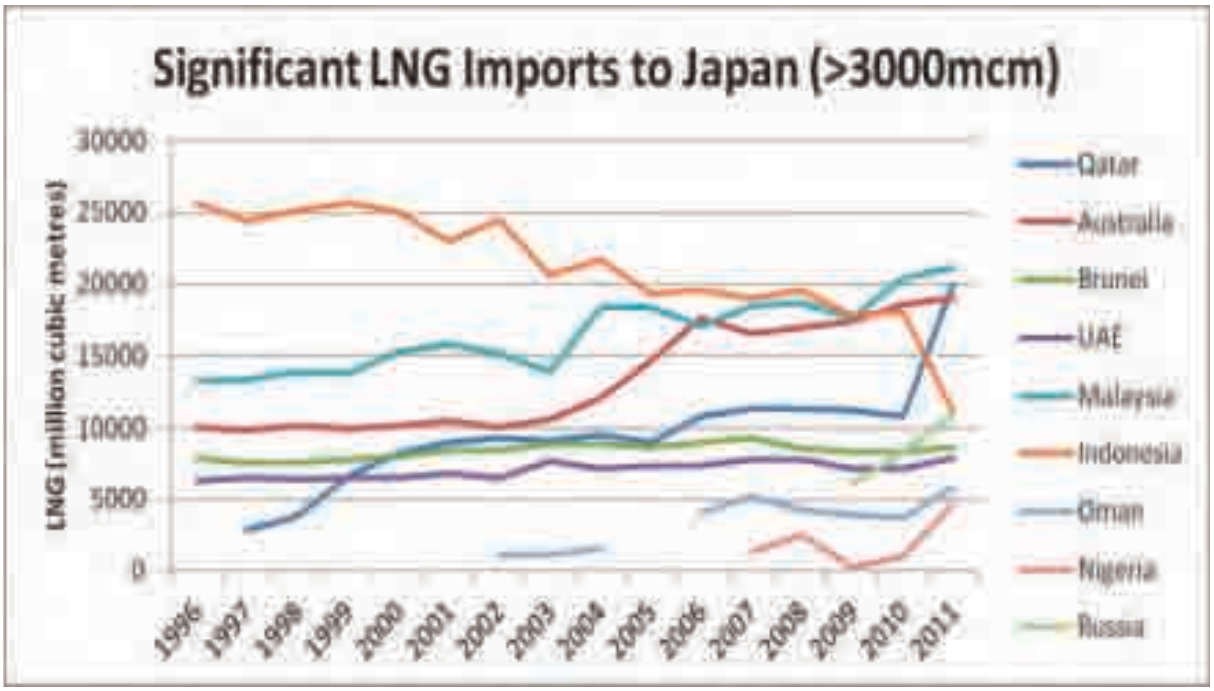
Duly Qatar began to supply Japan with LNG and continues to this day. Indeed, from the beginning in 1996 Qatar's importance to Japan has inexorably increased and in 2011 Japan became the second largest exporter of LNG to Japan behind Malaysia.

---

[13] For an thorough explanation of the machinations in the 1980s and 1990s regarding the Japanese-Qatari courtship see Kohei Hashimoto, Jareer Elass, and Stacy Eller, "Liquified Natural Gas from Qatar: The Qatargas Project," in *Geopolitics of Gas Working Paper Series* (Stanford: Baker Institute Energy Forum, Rice University, December 2004).

[14] Ibid., p.247.

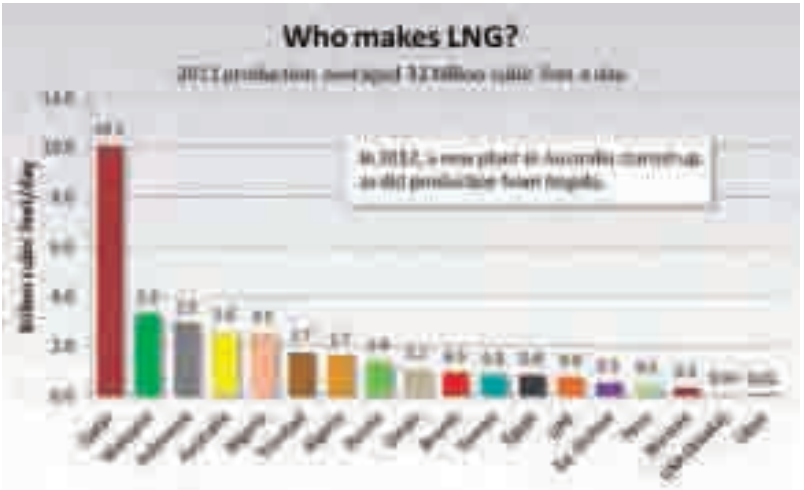
Figure 8 : Significant LNG Imports to Japan



Source: BP Statistical Review of World Energy[15]

Overall, Qatar's LNG production shot up in 2011 by over 25%, meaning that of the 87.7% increase in world LNG trade in 2011, Qatar accounted for virtually all of it.[16] As of 2011, Qatar was far and away the largest LNG producer in the world.

Figure 9 : Who makes LNG?



Source: The Oil Drum

---

[15] *Bp Statistical Review of World Energy*.

[16] Ibid., p.4.

Future Supply and Demand

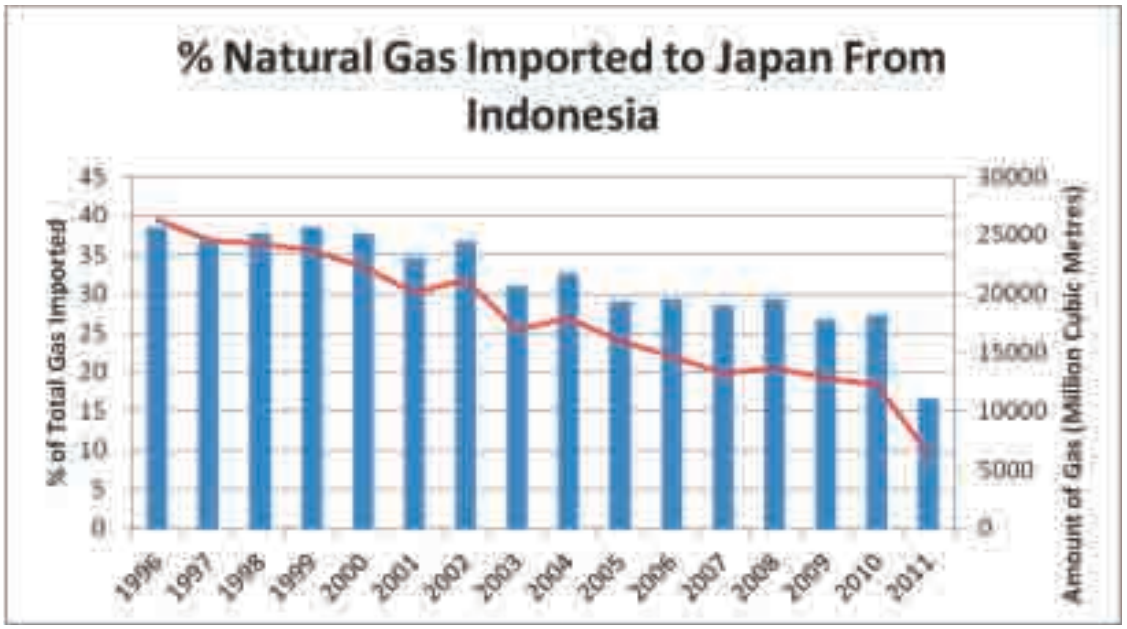
Thus far the Qatari-Japanese relationship has been ever more solidified by their continuing cooperation and trade, particularly in the area of LNG. However, there are several significant changes currently taking place both in the gas industry itself and elsewhere that will affect this key bi-lateral relationship in both positive and potentially negative ways. At its most basic level these new concerns can be split into concerns over supply and demand.

Supply

Indonesia

As noted, Indonesia has supplied Japan with a significant amount of its LNG in the past. Yet, as this graph simply but clearly demonstrates, the percentage of gas from Indonesia has been steadily declining for decades (red line). This in and of itself is not a cause for concern as it simply means that Japan is diversifying its sources of import. However, the absolute amount of gas has been dropping too and there are even predictions that Japan will not import any Indonesian gas by as early as 2014.[17] In a short period of time, this is a remarkable change and a large supply-side hole to fill.

Figure 10 : % of Natural Gas Imported to Japan From Indonesia



Source: BP Statistical Review of World Energy[18]

However, Japan's policy of diversifying its sources of import has been sensible. As of 2010 Japan had

---

[17] Stone, "Gas & Geopolitics: The Foreign Policy Implications of Energy Import Dependency", p.62.

[18] *Bp Statistical Review of World Energy*.

15 suppliers of LNG on whom to rely when the 15 becomes 14 as and when Indonesia stop supplying Japan. Moreover, as noted, Qatar's vast increase in 2010 and 2011 have occurred at a fortunate moment for Japan meaning that Qatar can to a large degree take up the slack left by Indonesia.[19] This does, however, diminish Qatar's ability to act as a major swing producer.

Shale Gas

The American gas market has had something of a rollercoaster time in recent years. From a state of almost self-sufficiency in the 1990s, imports increased as the 2000s progressed as domestic production declined.[20] Indeed, the market looked more and more promising for LNG exporters keen to fill the void and degasification terminals were duly constructed. Now, however, with America returning to its self-sufficiency and more, these terminals are either mothballed or turned around to be used for America's nascent LNG industry.

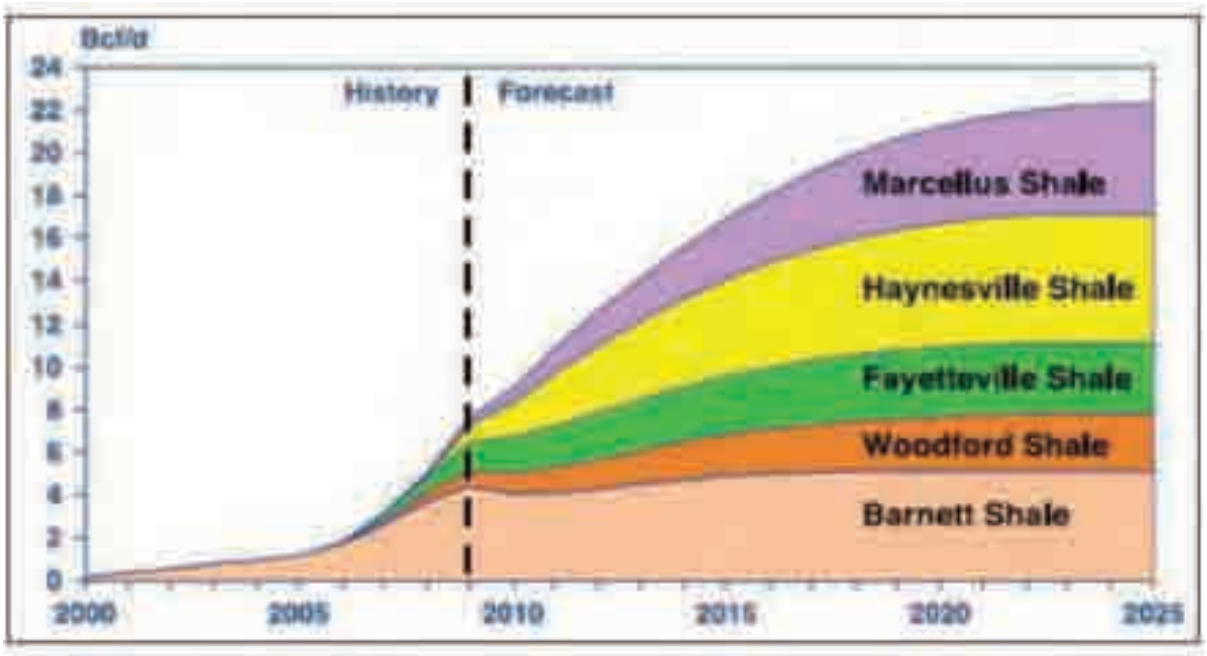
Indeed, the primary cause for this resurgence in American gas is thanks to advances in the area of unconventional gas procurement. The shale gas revolution has caused quite a stir. Deposits of gas in shale rock formations have been known for decades but the technology to feasibly and cost-effectively free the gas has not been available until recently. The processes involved are controversial. They involve pumping the most toxic of cocktails into rock formations to break down the rock. There have been many complaints that this 'fracking' process has led to water table pollution and mystery illnesses in areas with large concentrations of fracking. Nevertheless, America is leading the way thus far.

---

[19] Hirohide Hirai, "Lng Supply and Demand after the Great East Japan Earthquake and Japan-Russia Cooperation," in *The Fourth Japan-Russia Energy and Environment Dialogue* (Niigata, Japan15 November 2011), p.3.

[20] James Henderson, "The Potential Impact of North American Lng Export," (Oxford: The Oxford Energy Institute, October 2012), p.1.

Figure 11 : Predicted Shale Gas Production From Certain US Basins



Source: Paper at the International Petroleum Conference[21]

While there are many variables still to be settled, so far issues of price look to be encouraging. Longer term European prices typically hover around the $12/MMBtu mark, while in Asia the price has risen to $18/MMBtu recently. The US domestic spot price, however, in 2012 was a cheap $2/MMBtu. Quite clearly there is (potentially huge) opportunity here for arbitrage. Indeed, estimates from several of America's shale basins have calculated that US shale would need to sell in Japan and Korea for around $7.17/MMBtu to make a profit.[22] This is a highly competitive price considering that Japan has been paying nearly £18MMBtu in 2012.[23]

Already the fruits of the US shale revolution can be seen in Japan. In July 2012 a 20 year liquefaction agreement was signed with Osaka Gas and Chubu Electric, with plans to ramp up eventually to 13.2mtpa. Additionally, other Japanese consortiums (Mitsubishi and Mutsui) have signed agreements to participate in other US shale liquefaction projects including Sumitomo and Tokyo Gas signing for 2.3mtpa with a twenty year contract from Cove Point.

---

[21] Chau Tran et al., "International Lng Prospects 2011 and Beyond," in *International Petroleum Technology Conference* (Bangkok, Thailand2011), p.10.Graph taken from

[22] Ibid., p.9.

[23] In reality, the US price would doubtless be higher as cost overruns and delays sap the price, but this is nevertheless an interesting note.

Potentially, according the US Government estimates, America could corner around 19% or 38% of total LNG trade in the world, depending on whether one takes the low or the high estimate of US LNG production. Moreover, if every single project were to come online then theoretically America could corner more than 75% of the world's LNG trade.[24] However, none of this happens in isolation. As US LNG comes onto the market, it will interact with contracting and expanding supply and demand elsewhere. The price will consequently vary especially if the US market begins to dominate, flattening out the differences between the three major price hubs today (one price hub is broadly for US gas prices, one is European based, while another is based on a particular price cocktail in Japan).

Lastly, it must be noted that it is China's and not America's shale gas reserves that are thought to be the largest in the world. However, the shale fracking technology and expertise lies in America and China is far behind on this matter. While they could develop or steal the technology it is not that simple, as careful regulatory mechanisms are needed to ensure the safe and efficient production of shale. More and more Western firms are trying to help China in this endeavor, especially after China indicated its willingness to cooperate with Western majors, but making a dent in China's energy demands is still a long way off. However, if the industry does get to its feet to the degree that Chinese leaders demand, then it will radically alter the world energy market.

Australia

For some years now Australia has been positioning itself to become one of the leading players in the international gas market. Indeed, some suggest that Australia could overtake Qatar as the world's number one LNG producer by 2022, giving Qatar stiff competition particularly regarding East Asian suppliers.

Thus far, Australia has the world's 11th largest proved supplies. LNG has been going in Australia for decades, but has recently been ramped up. There are three major terminals presently making Australia the 4th largest supplier in the world at 7.9% of exports. Japan is already a key customer taking 73.4% of Australia's exports while China received 19.3%.[25] However, several new LNG projects are under construction. Indeed, six of the ten most expensive energy projects in the world are currently being completed in Australia for a total of $236 billion (GLNG $30BN, Queensland Curtis LNG $34bn,

[24] Henderson, "The Potential Impact of North American Lng Export," p.39.

[25] "Tankers on the Horizon: Australia's Coming Lng Boom," (London: The Economist Intelligence Unit, 2012).

Wheatstone $35bn, Australia Pacific LNG $37bn, Ichthys $43bn, Gorgon $57bn).[26]

Demand

Asian Demand

The issue is far more complex than simply filling the supply gap left by Indonesia for the gas industry, like all complex industries, is subject to market fluctuations, technological ramifications, and ebb and flow of international relations.

Japan finds itself situated in a highly competitive corner of the world. Since 1980 Asia has been experiencing 4.6% growth in annual energy demand; more than twice the average of the rest of the world.[27] Equally, the Asia Pacific region has been demanding LNG specifically at an increasing rate of 6.5% per year from 2007-2011.[28] While the Asia-Pacific has increased its domestic gas production 76% from 2000-2011, this is far from enough to cover the 500% increase in gas demand.[29] Lastly, when one considers that Asia currently accounts for only 30% of energy demand when its GDP per capita is 10% that of the OECD average, taking into account the growth of Asian economies (at least 5% per year), the demands that will emanate from this continent in the coming decades are on a clear trajectory. Moreover, given the acute pressure from climate change and the mitigation policies and procedures in place, LNG is ever more central as the viable fuel source given its relatively low carbon cost.

China currently obtains the vast majority (70%) of its energy in the form of coal, while gas accounts for only 4%; an energy mix not seen since the Industrial Revolution, as some commentators put it.[30] Existing pledges by the Chinese Government decree that gas should construe 10% of the energy demand by 2020. Even this relatively small percentage increase would call for a huge amount of LNG to be imported, for China's domestic reserves cannot nearly match such demand.

---

[26] Steve Hargreaves, "10 Most Expensive Energy Projects in the World," *CNN* (27 August 2012), http://money.cnn.com/gallery/news/economy/2012/08/27/expensive-energy-projects/index.html

[27] Ryoichi Komiyama, "Energy Outlook to 2035 in Asia and Its Pathways Towards a Low Carbon Energy System," in *World Energy Council* (The Institute of Energy Economics, Japan), p.1.

[28] Henderson, "The Potential Impact of North American Lng Export," p.2.

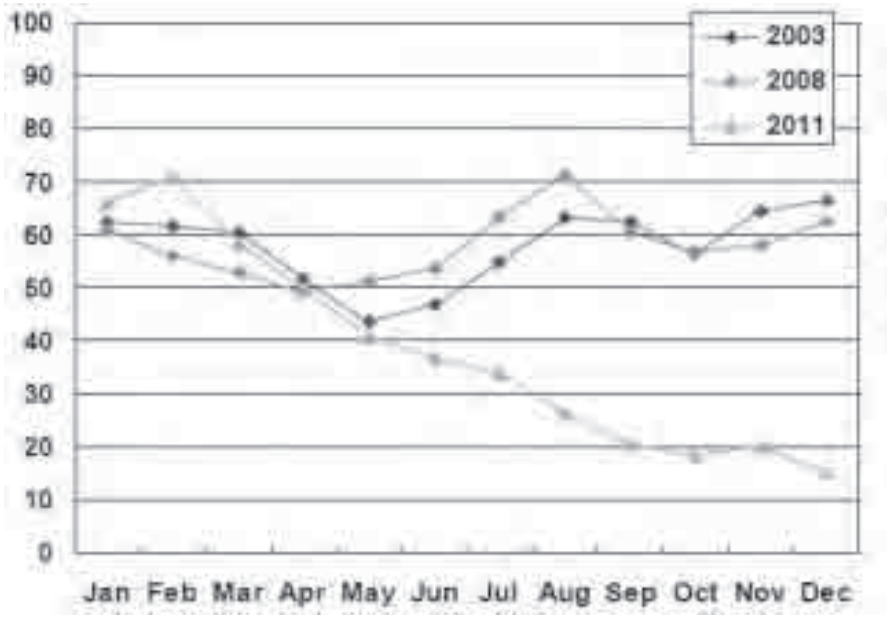
[29] Ibid., p.5.

[30] Tran et al., "International Lng Prospects 2011 and Beyond," p.6.

Fukashima

The Great East Japan Earthquake will have ramifications for years to come. Analyzing exactly what they will be is difficult. In the immediate aftermath there was an energy gap. The reactors themselves at Fukashima contributed 2.2% of Japan's total electricity demand and nuclear energy as a whole 33% and thus without this source, the gap needed to be filled.[31]

Figure 12 : Monthly Nuclear Load Factor in 2003, 2008 and 2011



Source: Statistics of Electricity, METI Website. Graph taken from Oxford Energy[32]

The preceding graph shows the consequent drop off in nuclear power as other affected stations were taken off line and others underwent rigorous testing. This energy gap was in part made up with LNG. The six months after the earthquake saw a year-on-year increase of 18% in LNG imports. Given that by February 2012 only two of the 50 Japanese Nuclear Power reactors were operational, this extra reliance is not going away.

Indeed, the confused government reaction with sporadic and at times unlawful closures of various nuclear power plants exacerbated the uncertainty and forced the local energy companies to revert ever more to LNG. Russia, Indonesia, and Qatar all increased their LNG supplies to Japan.

---

[31] Ibid., p.5.

[32] Akira Miyamoto, Chikako Ishiguro, and Mitshuiro Nakamura, "A Realistic Perspective on Japan's Lng Demand after Fukushima," (Oxford: The Oxford Institute for Energy Studies, June 2012), p.9.

The global down turn which was hitting Europe and America hard at the time meant that there was excess LNG that could be redirected to Japan. America's relatively sudden shale gas output also freed up spare LNG on the international 'spot market'. Indeed, as the Oxford Energy institute's report on this incident noted, the spot price was lower than normal contracted prices.[33] Another corollary of this whole incident is that new Japanese contracts, given the assumed greater demand in the future, are being set somewhat higher in price than previously.

Obviously, plans afoot before Fukashima to increase the numbers of nuclear power stations are, at best, on hold. The Government had announced that there were plans to increase the number of reactors by 41% by 2017 and 53% by 2030 with the construction of nine new reactors by 2020 and five by 2030.[34] However, today nuclear power is deeply unpopular in Japan and these considerations will need to be revised. Survey date quantifies the decisive and unsurprising decline in public disposition towards nuclear power.

Figure 13 : Survey Data on Public Perception of Nuclear Industry Pre and Post Fukashima

Date of Survey	2005.12 (*1)	2009.10 (*1)	2011.06 (*2)	2011.10 (*2)
Expand	35.1	39.6	5.0	2.1
Maintain status quo	30.2	18.8	24.4	23.2
Reduce and decommission	17.0	16.2	66.1	66.6
Don't know	7.7	5.4	4.5	8.2

*Note: *1 Published by the Public Relations Department of the Cabinet Office*

Survey taken from Oxford Report[35]

Conclusions

The supply and demand side issues are inextricably linked and no straight-line predictions can be made. Taking a moderate, nuanced view of the supply side, it is likely that America will enter the LNG market to a relatively significant degree in the next ten years. Simultaneously, the $100bn plus being invested in Australia as we speak will, even with predictable cost overruns, delays, and problems, inevitably dramatically increase Australia's LNG footprint, to say nothing of mooted growth in East African ventures or Russia's desire to enter the LNG market ever more.

---

[33] Ibid., p.18.

[34] "World Energy Perspective: Nuclear Energy One Year after Fukashima," (World Energy Council, 2011), p.13.

[35] Miyamoto, Ishiguro, and Nakamura, "A Realistic Perspective on Japan's Lng Demand after Fukushima," p.32.

In terms of demand, Japan's energy mix post the Great East Japan Earthquake will continue to be significantly skewed towards greater LNG demand. At the same time Korea and China, to name but two significantly growing economies, will doubtless increase demand too.

On the surface, therefore, there may be a joining of new demand and new supply though it seems likely that there may be something of a relative glut in supply. While this is potentially good for Japan, a lower price may not be good for Qatar.

Yet the relationship between the two countries is quite clearly strong. Though the elites that signed the original deals are retiring, there is an obvious push led by Japan though taken on by both sides to reinforce the link to a new generation. Qatari aid to Japan after the Earthquake, not to mention its LNG cargoes, was significant and reflected the strong elite-led desire to aid a country that played a key role in nurturing Qatar's key industry. Indeed, one must not forget the critical role of elite-led diplomacy in Qatar. Power is utterly centralized in the hands of but a few members of the elite.

The former, long-term energy Minister, Abdullah Al Attiyah, as noted, is one of the Emir's closest confidants and has been for decades. He was the architect on the Qatari side of the Japanese relationship regarding LNG. Though he retired from his position as Energy Minister some years ago, by virtue of his key importance as a key Emiri ally and his role in establishing Japanese relations, he – and not the current Energy Minister – nevertheless still leads Qatari missions to Japan today. Doubtless Al Attiyah was a strong influence behind the recent Qatar-Japan year event, seeking to augment ties between the two countries.

In many ways, given the way that Qatari politics functions, it is the fact that the event was organized and supported by both sides that is the key and not necessarily that the events themselves actually took place. Though exact figures are not available, it is not thought that many Qataris attended the Japanese events, yet as long as the younger members elite understand the importance of the relationship to key actors like the Emir and to Al Attiyah, as made clear by the elite's recent policy decisions, the relationship is set.

Bibliography

Abdulkarim, Rami "Natural Gas Security of Supply in Japan." *Shingetsu Electronic Journal of Japanese-Islamic Relations* 5, (March 2009).

Bp Statistical Review of World Energy: London : BP, June 2012.

Calabrese, John "Japan's New Energy Future and the Middle East." *The Middle East Institute* (11 June 2012), http://www.mei.edu/content/japan%E2%80%99s-new-energy-future-and-middle-east Access Date,

Hargreaves, Steve "10 Most Expensive Energy Projects in the World." *CNN* (27 August 2012), http://money.cnn.com/gallery/news/economy/2012/08/27/expensive-energy-projects/index.html Access Date, 16 October 2012.

Hashimoto, Kohei, Jareer Elass, and Stacy Eller "Liquified Natural Gas from Qatar: The Qatargas Project." *Geopolitics of Gas Working Paper Series*, Stanford: Baker Institute Energy Forum, Rice University, December 2004.

Henderson, James "The Potential Impact of North American Lng Export." Oxford: The Oxford Energy Institute, October 2012.

Hirai, Hirohide "Lng Supply and Demand after the Great East Japan Earthquake and Japan-Russia Cooperation." Presented at *The Fourth Japan-Russia Energy and Environment Dialogue*. Niigata, Japan, 15 November 2011.

"Japan." *Country Analysis Briefs*, Washington: Energy Information Administration, 4 June 2012.

Komiyama, Ryoichi "Energy Outlook to 2035 in Asia and Its Pathways Towards a Low Carbon Energy System." *World Energy Council*: The Institute of Energy Economics, Japan.

Lesbirel, Hayden S "Diversification and Energy Security Risks: The Japanese Case." *Japanese Journal of Political Science* 5, no. 1 (May 2004).

Miyamoto, Akira, Chikako Ishiguro, and Mitshuiro Nakamura "A Realistic Perspective on Japan's Lng Demand after Fukushima." Oxford: The Oxford Institute for Energy Studies, June 2012.

"Qatar: Background." *Country Background*, Washington DC: U.S. Energy Information Administration, January 2011.

Stone, Sam "Gas & Geopolitics: The Foreign Policy Implications of Energy Import Dependency." Stanford University, May 2010.

"Tankers on the Horizon: Australia's Coming Lng Boom." London: The Economist Intelligence Unit, 2012.

Tran, Chau, Bruce Cassidy, Marisa Pierce, Xiuli Wang, and Michael J Economies "International Lng Prospects 2011 and Beyond." Presented at *International Petroleum Technology Conference*.

Bangkok, Thailand, 2011.

"World Energy Perspective: Nuclear Energy One Year after Fukashima." World Energy Council, 2011.