2016 年 USACSA 赴兰大、云大与理大三校春晖学术交流团总结报告

美国华裔教授专家协会

一、概要

2016 年 6 月 7 日至 17 日，我们美国华裔教授学者协会(USCSA)春晖计划交流团一行 6 人在中国国家教育部、中国驻洛杉矶总领馆的全力支持下，在兰州大学、云南大学与大理大学各 级领导、三校国际合作与交流处领导和老师的悉心安排和帮助下，先后成功地对兰大(5 人交流团)、云大(5 人交流团)与理大(4 人交流团)进行了为期 3 至 10 天的学术交流和访问。其中参加三校交流者 3 人，两校(云大与兰大或理大)校交流者 2 人与一校(兰大)交流者 1 人。作为针对三校学科认真挑选的各领域的学术 带头人，各位专家学者和三校有关院校的领导和学术带头人进行了广泛学术交流，举办了一系列的学术专题讲座，将自己领域的最新研究动态和成果介绍给三校同仁，并且与 各院校学者之间展开了活跃的互动交流和商讨。春晖计划学者也和各院校认真探讨了在学术研究和教育培养方面今后可能合作的领域与项目。美国华裔教授学者协会与兰州大学签订了合作协议书。六位专家学者个人亦与所访问的学校在科研、教育和师资培养等方面达成了书面或口头合作协议或初步意向。通过这次春晖计划，我们对三校的科研方向和重点有了深入一步的了解，为今后合作和交流奠定了基础。各位学者表示 将与有关院校保持通畅的联络渠道和争取建立长期科研与教育合作伙伴关系。同时也将为三校和南加州地区各学术机构，包括南加州大学(USC)、加州大学各分校(如 UCLA, UCR, UCI)等院校在更多领域 的学术交流和合作牵线搭桥。访问期间春晖计划专家学者们严谨的学术态度、忘我的奉献精神和全力以赴的工作干劲也给所访问的学校留下了深刻的印象。

二、春晖计划代表团成员和交流项目

1).邢永刚教授 (美国南加州大学；访问三校)：(A) 地震断裂带共震破碎和震后愈合的观察、分析和模拟 (兰大资源环境学院：6-12-16)；(B) 应用断层导波确定汶川地震断裂结构 - 为灾后重建选址提供科学依据 (云大资源环境与地球科学学院：6-14-16)

2).黄胜和教授 (美国南加州大学；访问三校):(A) 宿主为靶向的抗病原微生物药物(兰大药学院：6-8-16);(B)感染组学共病两极(Sym-Pat)对偶问题解析：儿科抗感染治疗范式转换(兰大第一医院儿科感染系：6-12-16)；(C)HIV 感染与血脑屏障损伤研究进展 (理大微生物教研室:6-16-16)

3). 章华博士 (美国智泰科技公司；访问三校)：美国云计算发展现状和趋势 (兰大计算机网络中心：6-12-16)

4. 李百炼教授(美国加州大学河滨分校；访问兰大与云大两校)：(A) 数学建模及其在复杂生态系统研究中的应用 (兰大生命科学院：6-12-16)；(B) 未来新经济：人与自然和谐的生态发展 (云大：6-14-16) 。

5. 周歧发教授(南加州大学生物医学工程学系;访问云大与理大两校)：生物医学工程和医疗仪器 (云大工程系：6-15-16 下午座谈与讨论；参观一家医疗仪器公司， 并进行技术交流)

6). 黄南松教授(美国南加州大东亚研究中心；访问兰大)：对外语法教学语法体系之问题暨科学对外汉语教学语法之建立 (兰大文学院：6-12-16 上午与晚上两个讲座)

三、春晖计划代表团访问交流活动日程和内容

我们春晖计划代表团第一行程 5 人分别于 2016 年 6 月 7 日（黄胜和、章华）、9 日(郦永刚)和 11 日(李百炼、黄南松)陆续到达兰州机场。兰州大学安排专人在机场分批接机，随后送往兰州大学宾馆入住。黄胜和教授和章华博士于 6 月 8 日上午分别参观了兰大药学院与计算机网络中心。黄胜和教授应兰大药学院李红玉院长邀请 8 日下午作了题为”宿主为靶向的抗病原微生物药物” 的报告。李院长亲自主持报告会。6 月 9 日至 11 日为端午节休假日先期抵达兰州的团员自行安排活动。6 月 12 日黄胜和、郦永刚、章华、李百炼、黄南松等 5 个团员分别到对接单位(祥情见上)进行学术交流或讲座并恰谈合作项目(详情见下：各自总结）。兰大潘保田副校长于 6 月 12 日中午会见与宴请了代表团全体成员。6 月 13 日上午在离开兰州之前郦永刚教授与黄胜和教授代表美国华裔教授学者协会与兰州大学签订了合作协议书。6 月 13 日黄南松教授结束了春晖计划行程飞回北京。春晖计划代表团其他 4 位成员继续下一行程。

6 月 13 日我们春晖计划代表团第二行程 5 人分别从兰州（黄胜和、章华、郦永刚、李百炼） 和西安(周歧发） 两地飞往昆明天水机场。云南大学安排专人在机场分批接机，随后送往云南大学连云宾馆入住。6 月 14 日上午李百炼教授代表美国华裔教授学者协会访问团做了一个小时的题为“未来新经济：人与自然和谐的生态发展” 的报告。当天下午访问团教授专家分别到各院系作学术报告和交流。受到云大副校長，以及云大生态学和环境学院段昌群院长和云南省科技厅王学勤副厅长热诚招待。云大张力副校长于 6 月 14 日中午接见与宴请了代团全体成员。双方将进一步加强多方位合作。6 月 15 日李百炼教授结束了春晖计划行程飞往深圳。春晖计划代表团其他 4 位成员前往理大访问与交流。

春晖计划代表团第三行程 4 人(黄胜和、章华、郦永刚、周歧发)于 2016 年 6 月 15 日清晨乘飞机抵达大理大学。访问期间受到大理大学李小兵副校長，国际合作处房慧处长，以及各院系负责人热诚接待。在招待座谈会上李副校长和各院院长介绍了大理大学各学科现状，发展方向和对高水平人才的需求。座谈会后，欢迎仪式后，理大校方向每位学者赠送了礼品并合影留念。随后由校领导陪同共进欢迎午宴。当日下午各位专家分别安排在专业院系讲学和交流。

6 月 16 日下午章华、郦永刚、周歧发等三位专家飞回昆明并于 17 日飞往各自目的地结束春晖计划行程。黄胜和偕夫人吴春华女士(微生物专家，完全自费）分别于 6 月 16 日下午和 18 日上午顺便访问了理大微生物教研究室和昆明医科大学儿童医院并进行了卓有成效的学术交流。6 月 18 日下午此次春晖计划行程全部结束。

总之这次春晖计划是美国华裔教授学者协会在过去工作基础上一个学术和教育合作的良好续篇，是一个成功的旅程。此次春晖计划的成功，取决于中国国家教育部和中国驻洛杉矶总领馆的 全力支持，我们表示衷心的感谢。我们特别感谢洛杉矶总领馆教育组组长与参赞袁东博士和领事柴海鹰老师在百忙中抽出大量时间和精力处理有关的许多繁琐申请事务，感谢本会会长吴仲和教授、付会长郦永刚教授、理事会和会员们 对此计划的热情关注和帮助。我们也由衷地感谢三校各级领导、三校国际合作与交流处领导和老师的对交流程序的精心安排和食宿等各方面的关心照顾。总而言之，这次春晖交流计划的成功取决于各个部门领导和同事 的大力帮助，我们全体团员表示深深的谢意！

各位学者的交流汇报摘要如下：

郦永刚（南加州大学）**讲学交流活动简报如下：**

在兰州大学的讲学和交流活动

应兰大教育部重点实验室邀请，本人在兰州大学资源环境学院和地质科学与矿产资源学院作学术报告，题名为”地震断裂带共震破碎和震后愈合的观察、分析和模拟” 。约 2 小时的学术报告中，我介绍了在美国加里福尼亚州断裂带发现断层导波并应用导波研究断层深部精细结构和岩体物理属性的前沿研究工作，解析和模拟地震断裂带共震破碎和震后愈合的动力学过程，对地震周期性研究具有指导意义。专题报告内容丰富，博行好评。潘校长希望我作为特聘教授定期或不定期参与西部环境重点实验室的中国西北地震带研究项目。

在云南大学的讲学和交流活动

应云大资源环境与地球科学学院邀请，在该院作学术报告 ”应用断层导波确定汶川地震断裂结构 - 为灾后重建选址提供科学依据”。本人在 2008 年汶川 8 级大震后及时赶赴”中国地震局汶川地震科学考查指挥部” 受邀参与地震第一线科考工作。云大资源环境与地球科学学院院长谈树成教授， 地球物理系主任康国发教授等数十位教师和研究生在座。在报告会上，向云大地球物理系赠送我编写的新书(英文) ”(Seismic Imaging, Fault Damage and Heal, by Li 2014) ，供地震研究人员和研究生参考。该书荣获 2014 年度中国图书出版优秀图书奖。

在大理大学的讲学和交流活动

美国华裔教授学者协会”春辉计划”回国访问团受到大理大学李小兵副校長，国际合作处房慧处长，以及各院系负责人热诚接待。在招待座谈会上李副校长和各院院长介绍了大理大学各学科现状，发展方向和对高水平人才的需求。座谈会后，各位专家分别安排在专业院系讲学和交流。

黄胜和 （南加州大学）**讲学交流活动简报如下:**

团队活动总结:在过去 10 年中我组织和参加了本会的两次“春晖计划”学术交流活动。相对于上次新疆春晖四人行活动(单一学校)， 本次活动具有团体大(6 人) 、访问学校多(3 个加上顺访共 4 个) 、时间长(10 天) 、对接专业面广(文理两大专业六个不同领域) 、团队作用强(本会与兰大签订合作协议书)和效率高(连续作战平均 3 天访问一个学校) 等 6 大**优点**。本次活动中兰大国际处工作最出色、与本会互动最佳、对接专业最到位和交流成果最卓著。**缺点**:(A) 准备时间仓促;(B) 本会作用发挥欠佳;(C) 在与云大和理大专业对接上相对薄弱。**建议**:(A) 尽早计划提前准备(如春晖计划工作不强的学校需要更多时间作准备);(B) 进一步加强团员多元化(如增加院士级以上和年轻专家参加); (C) 更充分发挥本会组织作用(如成立专门工作小组) 。

个人活动总结:(1)**兰大活动**:我分别于 6 月 8 日与 12 日应邀访问了兰大药学院与第一医院并作了题为宿主为靶向的抗病原微生物药物(药学院)和感染组学共病两极(Sym-Pat)对偶问题解析(第一医院)的两个学术报告。我同第一医院儿科感染系的医护人员座谈、讨论与分析了若干诊断不明的疑难病例并签订了协助从感染组学上解决这些疑难问题的协议书。(2)**云大活动**:6 月 14 日上午参加李百炼教授代表本团所做的关于生态发展与未来新经济的报告。当天下午随同我多年的合作伙伴李百炼教授(10 年前春晖活动结缘)与云大生态学和环境学院段昌群院长和他的科研团队座谈和交流。随后我与周歧发教授一起参观了一家医疗仪器公司。(3) **理大活动**:在理大李小兵副校長与国际合作处房慧处长所主持的座谈会上,李副校长和各院院长介绍了大理大学各学科现状， 发展方向和对高水平人才的需求。我代表本团发言介绍了美国华裔教授学者协会的专业特点与人才优势。座谈会后，我到理大儿科医院参观并与该院科研教学骨干座谈和讨论儿科教育体制问题。我介绍了美

国儿科教育特点。该院领导提出了儿科人才培养意向。6 月 16 日上午我应邀到理大微生物 教研室介绍了 HIV 感染与血脑屏障损伤研究进展并与该室达成了合作意向。(4) **顺访昆明 医科大学儿童医院**:我偕夫人吴春华女士(微生物专家,完全自费)于 6 月 18 日上午顺便访问了昆明医大儿童医院。在该院院长和刚从我实验室学成回国的博士后李莉医生陪同下参观了新建院区并就该院新建儿科研究所达成了合作意向。

李百炼(加州大学河滨分校) **讲学交流活动简报如下:**

2016 年 6 月 11 号抵达兰州大学，12 号上午在生命科学院做了 2 个小时的学术报告，题目是“数学建模及其在复杂生态系统研究中的应用”，下午与兰大大气科学院进行了学术交流并指导他们的论文写作。13 号下午抵达云南大学，14 号上午代表我们访问团做了一个小时的题为“未来新经济: 人与自然和谐的生态发展”，下午与云大生态学科的教师进行了学术交流并就高原湖泊生态学方面的合作达成协议。

黄南松（南加州大学）**讲学交流活动简报如下:**

本人于 2016 年 6 月 11-13 日同美国华裔教授协会的其他 4 名教授一起赴兰州大学进行“春晖计划”学术交流活动。现将具体情况报告如下：1. 为了保证学术报告适应兰州大学文学院的需要，早在 4 月 20 日本人即将若干学术报告的题目及简要介绍发给兰大文学院，由他们选定报告的题目。2. 6 月 12 日上午同兰州大学文学院及汉语国际教育专业负责人探讨今后合作的可能性，包括本人将来继续去兰大进行学术交流、协助邀请兰大教师赴美参加学术会议和进行学术访问。下午，同兰大汉语国际教育（硕士）专业的骨干教师就该专业的培养和发展问题进行座谈。3. 6 月 12 日晚上，举行“对外语法教学语法体系之问题暨科学对外汉语教学语法之建立”学术报告，参加者包括兰大文学院的教师、研究生及高年级学生共计 200 余人。4. 6 月 13 日上午，应约同兰大国际交流处长就如何利用海外资源发展兰大人文社会科学的问题进行了长达 2 小时的交流。因此，可以说，此次活动取得了很好的效果。在此，要感谢美国华裔教授协会前会长黄胜和教授的精心组织和驻洛杉矶领馆教育组的大力支持。

章华（美国智泰科技公司）**讲学交流活动简报如下:**

这次参加教育部赞助和洛杉矶总领馆倡导的春晖计划，十分荣幸，收获不少。我 6 月 6 号启程，7 号到达兰州大学，8 号就开始了学术交流。和我对接的是兰州大学计算机网络中心。我给网络中心做了“美国云计算发展现状和趋势”的讲座，为网络中心的现有工作和未来发展提出了建设性意见。我们草签了联合申请教育部春晖项目的协议。6 月 13 号离开兰州大学，我们到达昆明，对云南大学进行学术交流。我和云南大学软件学院的代表进行了比较深入交谈，对将来的合作提出了建设性的建议。6 月 14 号到达大理大学。我和大理大学数学和计算机系老师们进行了 2 个多小时的学术交流，并参观了机房和展览馆。6 月 16 号我们结束了这次学术访问。17 号离开昆明。

这次学术交流活动，前后差不多 10 天，行程紧凑，富有成效。认识了高校新朋友，也进一步熟悉了国内的学术圈。我的领域主要是云计算，大数据，物联网等。我把美国和西方在这方面的新动向，新知识和新技术向国内同行进行了汇报。建议：今后联系学校及对接单位应该由教授本人直接联系为好，这样便于后面的对接。这次到云南大学就出现一些问题。建议教授协会主要负责组织工作并提供可能的对接信息。

周歧发（南加州大学）**讲学交流活动简报如下:**

这次本人有幸参加了有黄教授组织的 2016 年六月的教育部春晖教授访问团。我是加入了云南大学和大理大学的活动。在云南大学， 有科技处安排了一个报告会， 我们全体成员

都参加了。 下午我们分组讨论， 我与黄教授一起参观了一位物理系老师合办的一家医疗仪器方面的公司， 并对我们的技术进行了分享。 双方回来后进一步讨论并商讨合作事宜。随后我们一起去了大理大学参观和访问， 下午主管外事的校长介绍了该校的有关情况。 第二天上午我们分组去了工程学院开会。与有关老师分享了有关我们的研究和教学工作。发现这个学校是一所刚成立的大学， 老师大部分是硕士毕业。 我们鼓励他们考在职博士研究生， 来提高研究水平。 目前已有一到二位老师有保持联系想进一步保持合作。通过这次访问收获还是很大， 一是看到了国内变化很大，另一方面让我们看到了西部人才还是缺乏，需要多加强合作。
建议是以后去之前与国内的有关单位联系好，将可能的合作单位和个人找到， 这样效果会更好！

Subsurface Damage Zone of the Calico Fault Viewed by Fault-Zone Trapped Waves from Teleseismic Earthquakes

Yong-Gang Li

Department of Earth Science, University of Southern California

Abstract. The fault-zone trapped waves (FZTWs) are first-time observed at a square seismic array consisting of 40 intermediate-period stations and 60 short-period stations deployed in a 1.5 km by 5.5 km grid adjacent to the Calico Fault (CF) in Mojave Desert, California for teleseismic earthquakes. In the previous study, FZTWs generated by explosions and local earthquakes, traveltimes inverse and InSAR observations have imaged a low-velocity compliant zone along the CF within which seismic velocities are reduced to the maximum of ~40-50% in the center of a ~1-km-wide fault core zone extending to at least 5-6 km depth (Cochran, et al., 2009). In order to obtain further constraints on the deep portion of the CF below the depth coverage of local earthquakes, we innovatively use the FZTWs recorded at this seismic array atop the CF compliant zone for teleseismic earthquakes. This type of FZTWs has not been used before, but they appear to have great promise for providing unprecedented constrains of the depth extension of fault-zone damage structure because these FZTWs arise from seismic waves incident at the fault bottom at deep level. We examined the data from 26 M≥6 teleseismic earthquakes occurring at distances of ~40^o-80^o great circles to the Calico array site during 2006. We identify this type of FZTWs at dominant frequencies 0.5-2 Hz, which show much larger amplitudes and longer wavetrains (~6-8s) starting ~5-s after the first-arrival P-waves at stations located within the ~1-km-wide CF compliant zone than those registered at farther stations out of the zone for teleseismic earthquakes. We synthesize these FZTWs formed by S-waves converted from the first-arrival P waves at the Moho (~30-km depth beneath Mojave Desert) and entering the bottom of the CF compliant zone. Combined with the previous velocity model of the CF, the results from observations and simulations of these FZTWs indicate that the compliant zone with velocity reduction of 40-50% along the CF likely extends to ~8-km depth.

1. Introduction

The Calico fault is located midway between the 1992 Mw7.3 Landers and 1999 Mw7.1 Hector Mine ruptures where fault-zone trapped waves (FZTWs) generated by explosions and aftershocks were observed (Li et al., 1994, 2000, 2002, 2003). Some geodetic estimates suggest that the CF has as much as 7 mm/yr dextral slip rates within the eastern California shear zone (Peltzer et al., 2001). Recent trench data indicate that the last earthquake occurred on it at least several hundred years ago (Ganev et al., 2008). Coseismic interferograms for both the Landers and Hector Mine earthquakes showed strain localized on the CF and other nearby faults (Fialko et al., 2002; Fialko, 2004). Line-of-sight displacements with amplitudes of a few centimeters and wavelengths of a few kilometers are clearly associated with the CF (Fig. 1a). Observed interferometric synthetic aperture radar (InSAR) anomalies for the CF are best explained with a 1–2-km-wide zone around the fault with a shear modulus reduced by ~50% extending to at least 5 km depth.

Cochran, et al. (2009) conducted a seismic investigation of the compliant zone along the Calico Fault (CF) in the Mojave Desert in 2006 (Fig. 1a, b). Using seismic travel times, FZTWs generated by explosions and local earthquakes as well as interferometric Synthetic Aperture Radar observations, they document seismic velocities within the ~1-km-wide CF compliant core zone reduced by up to ~40%, and shear moduli reduced by up to 60% compared with wall-rock (Fig. 1c). This low-velocity compliant zone extends to at least 5-6 km depth. The velocity reductions observed along the CF compliant zone are likely due to the cumulative mechanical damage from past earthquake ruptures on it and also indicate that faults can affect rock properties at substantial distances from the primary fault slip surfaces and throughout much of the seismogenic zone. This result has implications for the portion of energy expended during rupture to drive cracking and the development of fault systems. Permanent damage

zones may, thus, play a critical role in development and dynamics of faults, intensity of ground motion, and earthquake hazards along the low-velocity fault zone resulting from its waveguide effect. However, the deep portion of the fault damage zone is still less constrained by the FZTWs used in the previous study because those FZTWs generated by near-surface explosions and local earthquakes mostly travel along the fault zone at shallow depths.

Recently, we identified a new type of FZTWs in seismograms recorded in the 2006 experiment at the dense square array atop the Calico Fault for teleseismic earthquakes. Since these FZTWs arise from S-wave converted from P-wave at the Moho discontinuity and sub-vertically incident at the bottom of CF compliant zone. These FZTWs are capable to provide unprecedented constraints on the depth extension of fault damage structure beneath the depth that those FZTWs generated by explosions and local earthquakes can approach. In this article, we demonstrate the new type of FZTWs recorded at the dense array atop the CF from the teleseismic earthquakes (Fig. 1d) and use these FZTWs to image the subsurface fault rock damage structure at deep depth.

2. The Data and Waveform Analyses

Cochran et al. (2009) installed a dense array of 40 intermediate-period (40T) stations and 60 short-period (L22) stations in a 1.5 km × 5.5 km square adjacent to the Calico fault (CF) to record three shots detonated within and out of the Calico Fault zone, and then to record earthquakes for six months starting from June in 2006. This square array consists of 4 seismic lines (Line A, Line B, Line C and Line D in Fig. 1a) nearly perpendicularly across the CF with 1-km line spacing. The geometry of the array and station intervals is shown in Fig. 1b. Each station has a short-period sensor (L22 2Hz) and/or an intermediate-period sensor (40T 0.025Hz) buried in a 0.5-1-m deep dig hole with three components in the vertical, parallel and perpendicular to the fault strike. The seismometers worked in continuous mode at 100 samples per second, and were synchronized by internal GPS clocks. The data recorded at four cross-fault lines A, B, C and D for 3 explosions and 5 local on-fault earthquakes have been used for FZTWs analysis, and 20 local and 8 teleseismic earthquakes have been used for travel time tomography (Cochran et al., 2009), showing a ~1-km-wide compliant core zone along the CF extending to at least 5-6 km depth, within which seismic velocities are reduced by up to ~40% and shear moduli reduced by up to 60% compared to wall-rock (Fig. 1c).

In the present study, we examine the data recorded at 4 Lines A, B, C and D for 72 teleseismic earthquakes with magnitudes M≥6 and occurring at distances of ~40^o-80^o great circles (approximate 5,000-11,000 km) to the Calico seismic array site during June-November of 2006 (Fig. 1d). We use 26 teleseismic earthquakes at depths deeper than 28 km, for which we identify FZTWs closely after the first P-arrivals in teleseisms recorded at 4 cross-fault Lines A, B, C and D (Fig. 1b). These FZTWs likely arise from S-waves that are converted from the first-arrival P-waves at the Moho discontinuity (at ~30-km depth beneath Mojave Desert) and entering the bottom of the low-velocity compliant zone along the CF. Seismograms recorded for shallow teleseismic earthquakes are in general dominated by strong surface waves so that they are not suitable used in this study. In the present article, we show the FZTWs identified in teleseisms from 7 teleseismic earthquakes which locations and magnitudes are listed in Table 1.

Table 1. Teleseismic earthquakes and local earthquakes recorded at the seismic array atop the Calico fault, which waveforms are shown in this article

Event ID	Type	Magnitude	Latitude	Longitude	Depth (km)	Distance
TE162	Teleseism	6.3	33.130	131.150	140.0	72.2°
TE244	Teleseism	6.7	-6.780	155.540	38.0	69.5°
TE271	Teleseism	6.9	-16.613	-172.035	28.0	73.5°
TE272	Teleseism	6.1	10.910	-61.650	53.0	55.0°
TE276	Teleseism	6.3	-18.860	169.090	161.0	65.7°
TE288	Teleseism	6.7	19.878	-155.935	38.9	37.8°
TE317	Teleseism	6.8	-26.038	-63.243	547.5	78.9°

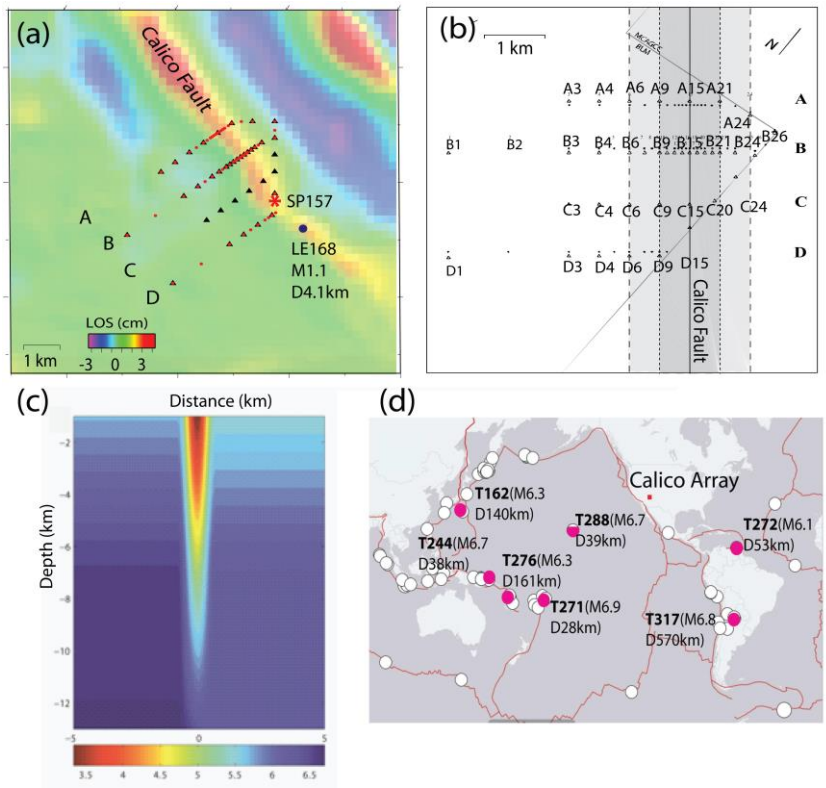


Fig. 1 (a) Map shows locations of 40 (40T, 0.025Hz) stations (black triangles) and 60 (L22, 2Hz) stations (red circles) along 4 seismic lines A, B, C and D deployed across the Calico Fault in 2006 (Cochran et al., 2009).. Colors in high-pass-filtered co-seismic interferogram from the 10/16/1999 Hector Mine earthquake during 01/13-10/20/1999 denote variations in line of sight (LOS) displacements (Fialko et al., 2002). **(b)** Map shows locations of intermediate- (triangles) and short-period (dots) stations along seismic lines A, B, C and D deployed across the CF compliant zone (marked by grey color) with the 1-km-wide fault core zone (dark grey). Station spacings are not even, but with denser stations within the compliant zone. **(c)** Model of P-wave low-velocity zone along Calico fault with 40% reduction in velocity obtained by Cochran et al. (2009). The fault compliant zone is ~1.5 km wide at the surface. The lateral velocity profile across the fault is approximated as a Hanning taper, and the velocity reduction tapers linearly to zero between 0 and 10 km depth. The damage zone is modeled to extend to 10 km depth, but with small relative velocity reductions below ~5-6 km. **(d)** Map shows locations of 71 teleseismic earthquakes (white circles) with M≥6 recorded at 4 seismic lines in the square array atop the CF. Red circles marked by labels with T followed by Julian date, M followed by magnitude and D followed by depth, denote seven teleseismic earthquakes for which FZTWs are identified in teleseisms and shown in this article.

Fig. 2a shows seismograms recorded at five 40T stations of seismic Line D across the CF for a M6.7 teleseismic earthquake (TE288 in Table 1) occurring at 39-km depth beneath Hawaii Islands and ~4200 km (37.8^{o}) west from the CF seismic array site in Southern California (see Table 1 and Fig. 1d). Because of the low level of noise at the array site, the primary P-wave and S-wave are visible in teleseisms while strong surface waves appear after 1000-s from the event origin time. We apply a band-pass (0.5-2.0 Hz) filter on teleseisms to suppress the surface waves, and cut seismograms from 430-s to 470-s, including the first-arrival P-wave (P') and the wavetrain closely following it. Fig. 2b exhibits the band-pass filtered seismograms at seismic Line D. We observe a wavetrain with large amplitude and ~8-s duration, in which the waveform amplitudes are twice above the background level, staring from ~5.5 s after the first-arrival P' at stations D9 and D15 located within the CF compliant zone (see Fig. 1b). We also observe this wavetrain at stations A9, A15 and A21 of Line A, and at stations C9, C15 and C20 of Line C located within the ~1-km wide CF complaint core zone for this teleseismic earthquake (Fig. 2c and Fig. 2d). In contrast, such waveforms do not appear at stations located out of the CF complaint zone. We interpret these long wavetrains with large amplitudes as being fault-zone trapped waves (FZTWs) formed by S-waves that are converted from first-arrival P waves at the Moho discontinuity (~30-km depth beneath

Mojave Desert) and sub-vertically enter the bottom of the low-velocity CF compliant zone at certain depth. We call P' for the first-arrival refracted P-wave and S' for the S-wave converted from P at the Moho discontinuity. The time difference between P' and S' is ~5.5 s Corresponding to the distance of waves traveling sub-vertically between the Moho at depth of ~30- km and seismic array at the surface. When the S' wave enters the bottom of the vertical fault zone at certain depth, S-type FZTWs are produced due to constructive interference of multiple reflected waves at the boundary between the low-velocity compliant zone and high-velocity surrounding rocks, and recorded at stations located within the CF compliant zone. We notice P-type FZTWs in the P coda before the S' in perpendicular-component seismograms recorded at stations located within the CF compliant zone for this teleseismic earthquake from which the ray path is nearly perpendicular to the CF strike.

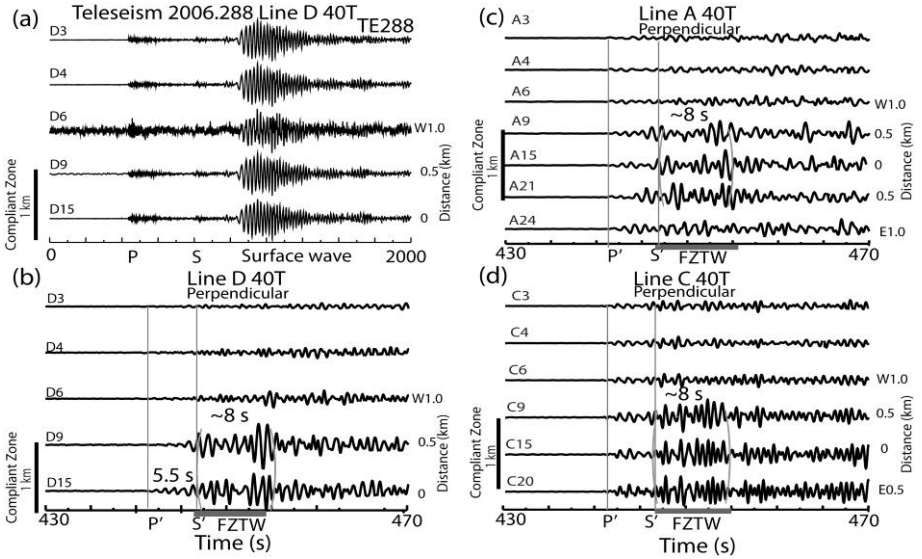


Fig. 2 (a) Seismograms in the fault-perpendicular component recorded at six 40T stations along Line D for a M6.7 teleseismic earthquake (TE288 in Table 1). Its location is shown by a red circle labeled with T288 in Fig. 1d. Station D15 is located on the fault trace. Station names and distance with respect to D15, and event date are shown in plot. Surface waves are dominant in raw seismograms, but primary P- and S-waves are visible too. The time scale in plot is 0-2000 s. **(b)** Band-pass (0.5-2.0 Hz) filtered seismograms recorded at Line D show prominent FZTWs (in brackets) with large amplitudes and long wavetrains (~8-s post-S duration marked by a horizontal bar) at stations D9 and D15 located within the ~1-km-wide compliant core zone (denoted by a vertical black bar). Note that the time scale in plot is 40 s starting before the first-P arrival. Seismograms are plotted using a fixed amplitude scale for all stations in each panel. P' and S' are P-waves and P-S converted waves from the Moho discontinuity between the lower crust and the upper mantle. **(c)** and **(d)** Same as in (b) but for seismograms at Line A and Line C, respectively. Stations A15 and C15 are located on the fault trace. Prominent FZTWs with large amplitudes and ~8-s post-S duration recorded at stations between A9 and A21, C9 and C20 located within the ~1-km-wide compliant core zone.

Fig. 3a shows three-component seismograms recorded at 15 stations equipped with intermediate-period 40T sensors and 25 stations equipped with short-period L22 sensors for a M6.3 teleseismic earthquake (TE162 in Table 1) occurring at 140-km depth in Honshu Prefecture of Japan, approximately 9500-km (72.2^{o}) west of the CF array site in California (see Fig. 1d). The first-arrival P' waves with its coda are dominant around 810-815-s on seismograms in vertical component, weaker in perpendicular-component, but weakest in parallel-component, because the waves come to surface stations near-vertically and are perpendicular to the CF strike. We observe large-amplitude and long-duration (~7-s) wavetrains arriving about 5-s after the P' in two horizontal-component teleseisms at stations between B9 and B21 located within the ~1-km-wide CF compliant core zone. In contrast, shorter wavetrains with much lower amplitudes are registered at stations out of the compliant zone, indicating abundance of seismic energy trapped within the CF compliant zone. We interpret that the observed wavetrains with long duration and large amplitudes in horizontal-component teleseisms are S-type FZTWs arising from

the S waves converted from P waves at the Moho (~30-km depth) and nearly vertically entering the bottom of the low-velocity compliant zone along the CF at the certain depth. These wavetrains are similar to the FZTWs generated by explosions and local earthquakes, and recorded at the same stations atop the CF (Cochran et al., 2009), but they are recorded for teleseismic earthquakes.

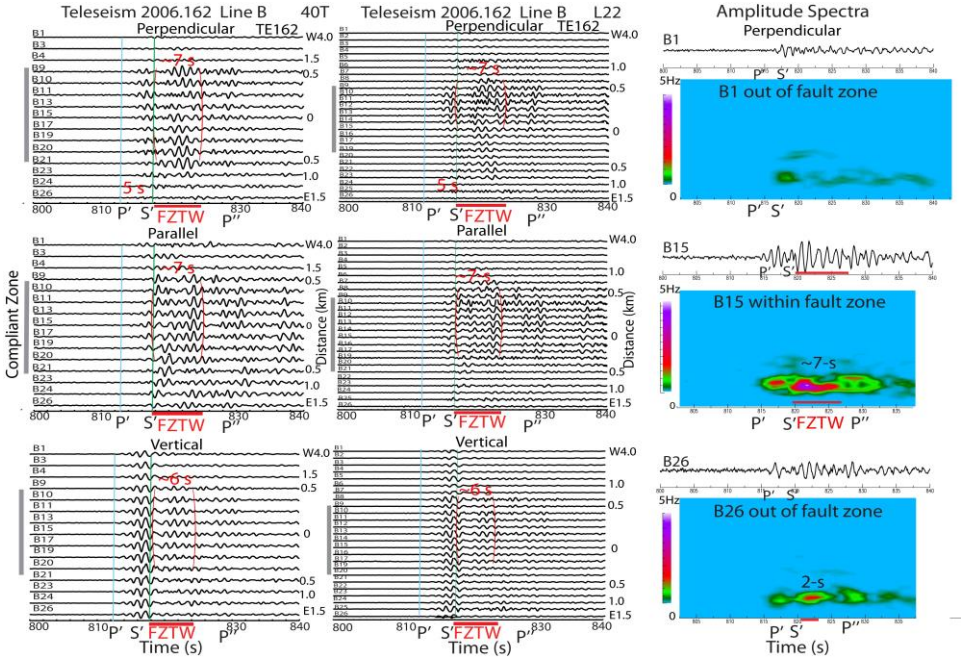


Fig.3a Left: Three-component seismograms recorded at fifteen intermediate-period 40T stations and twenty-six L22 2-Hz stations along Line B for a M6.3 teleseismic earthquake (TE162 in Table 1 and Fig. 1d). Station B15 is located on the fault trace. Seismograms have been band-pass (0.5-2 Hz) filtered and are plotted using a fixed amplitude scales for all trace in each plot, showing prominent FZTWs with ~7-s post-S duration (in red brackets) in two horizontal seismograms recorded at stations B9–B21 located within a ~1-km-wide compliant zone (marked by a vertical grey bar). P' and S' are P-waves and P-S converted waves at the Moho discontinuity. **Right:** Normalized spectral amplitude contours of teleseisms show large-amplitude (red color) FZTWs at 1-3 Hz with ~7-s duration at on-fault station B15, but lower amplitude with shorter duration (<2-s) at stations B1 and B26 out of the compliant zone. The same amplitude scale is used for three plots. P" denotes multiple P-waves between the surface and Moho.

We notice (1) the large-amplitude FZTWs following S'-waves on two horizontal-component seismograms at stations between B9 and B17 equipped with short-period 2-Hz sensors, indicating that the more seismic energy at higher frequency is trapped within a narrower (~500-600-m wide) fault core zone; (2) the long-duration wavetrains of FZTWs including multiple peaks (starting from P") like ripple, probably due to waves rebounded between the free surface and Moho; (3) the wavetrains in P coda immediately following the P' in vertical- and perpendicular-component seismograms likely being P-type FZTWs arising from P-waves entering at the bottom of compliant zone. In this article, we are interested in the S-type FZTWs following the S' converted from P at the Moho.

Fig. 3b shows this type of FZTWs with ~7-s durations following S' recorded at stations of seismic Lines A, B and C equipped with intermediate-period 40-T sensors located within the CF compliant zone for teleseismic earthquake TE162. These S-type FZTWs are dominant in horizontal-component seismograms, but not clear in vertical-component seismograms. The durations of FZTWs recorded for teleseismic earthquakes are longer than those generated by local earthquakes occurring near the seismic array atop the CF (Cochran et al., 2009; also see Fig. 9 in this article). Observations of these FZTWs suggest the existence of a 1--km-wide remarkable waveguide along the CF compliant zone likely extending from the ground surface through seismogenic depths.

We analyze P-to-S converted waves observed for the Moho discontinuity in teleseisms using the receiver function to obtain valuable constraints on the subsurface structure of the CF compliant zone

beneath Mojave Desert. The receiver function analysis is helpful to equalize near-source effects in three-component seismograms and isolate the receiver effects from the observed waveforms. We compute receiver functions by waveform deconvolution between the radial-component and vertical-component teleseisms for isolating the local response. Fig. 3c shows receiver functions computed for teleseisms recorded at five stations of Line B for teleseismic earthquake TE162. We interpret that the peak amplitudes of first-arrival P-wave (P') and P-to-S converted wave (S') from the Moho discontinuity in receiver functions appear at 0 s and ~5 s, respectively, as well as the multiple-phase P" at 14-15 s, according to the Moho depth of ~30-km and Vp in the upper crust beneath Mojave Desert is in average ~6-km/s. The large-amplitudes between ~5 s and 12 s in receiver functions at stations (B13, B14 and B15) located within the CF compliant zone likely correspond S-type FZTWs arising from S' wave entering at the bottom of the CF compliant zone at certain depth. In contrast, much lower amplitudes following S' in receiver functions at stations (B03 and B24) located outside the compliant zone. We note that high amplitudes following the P'-arrivals at stations (B13, B14 and B15) located within the CF compliant zone are likely P-type FZTWs. In contrast, there is no such amplitude peak in P coda at stations (B03 and B24) out of the CF compliant zone. Receiver functions suggest the existence of subsurface low-velocity structure along the CF compliant zone.

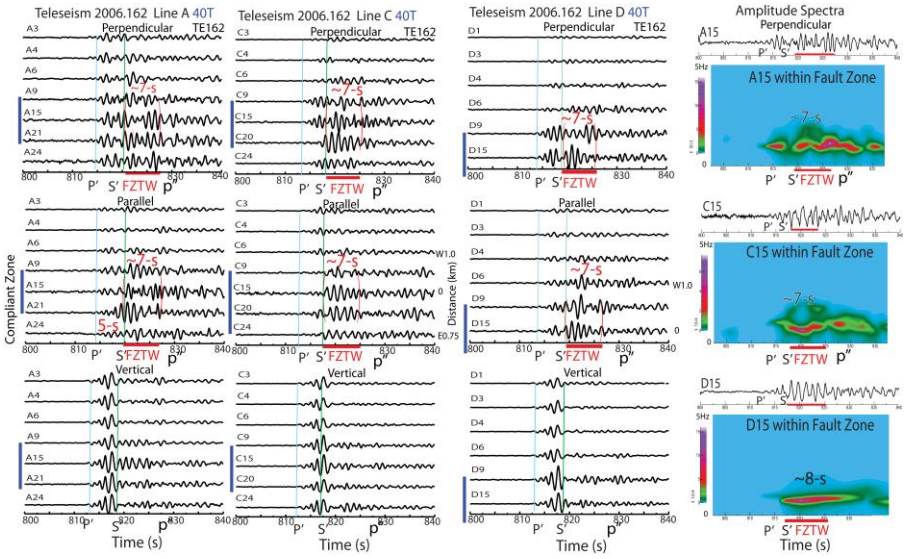


Fig.3b Left: Same as in Fig.4a, but showing teleseisms recorded at 40T stations along Lines A, C and D. Prominent FZTWs with large amplitudes and long post-S durations (~7-s) at stations located within the ~1-km wide compliant zone. **Right:** Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs with ~7-s duration at on-fault stations A15, C15 and D15. P" denotes multiple P-waves between the surface and Moho.

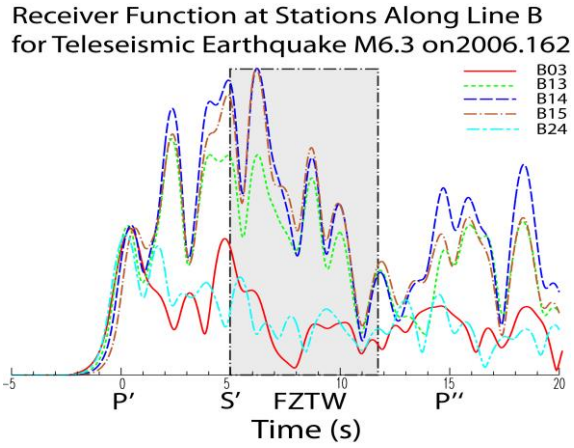


Fig.3c Receiver functions computed by deconvolution between the radial- and vertical-component seismograms recorded at 5 stations of Line B. P'-arrivals are aligned at time 0 s. P-to-S converted waves (S') arrive at ~5 s. The S-type fault-zone trapped waves (FZTWs) following S' show large amplitudes and ~7-s duration (in grey box) at stations (B-13, B14 and B15) located within the CF compliant zone while much lower amplitudes at stations (B03 and B24) outside the compliant zone. The multiple P" arrives at ~14-15 s. To emphasize the converted waves that reverberate between the surface and Moho, computed receive function curves have been smoothed to relieve the ripples from multiple fine layers. Note large amplitudes following P' and P" related to the P-type FZTWs.

In the next example, Fig. 4 exhibits three-component seismograms recorded at four lines A, B, C and D for a M6.3 teleseismic earthquake (TE276 in Table 1) occurring at 161-km depth beneath Solomon Islands, approximately 10,000 km (65.7^{o}) west of the CF seismic array in California (see Fig. 1d). We observe consistent large-amplitude wavetrains with long wavetrains (~7-s) after the P-S converted S'-wave at ~5-s after the first P-arrive, in two horizontal-component seismograms and spectral contours at stations located within the ~1-km-wide CF compliant core zone, but not at stations out of the compliant zone. These wavetrains are weaker in vertical-component teleseisms than those in horizontal components because waves from this deep teleseismic event propagate nearly-vertically from the Moho at ~30-km depth beneath the seismic array.

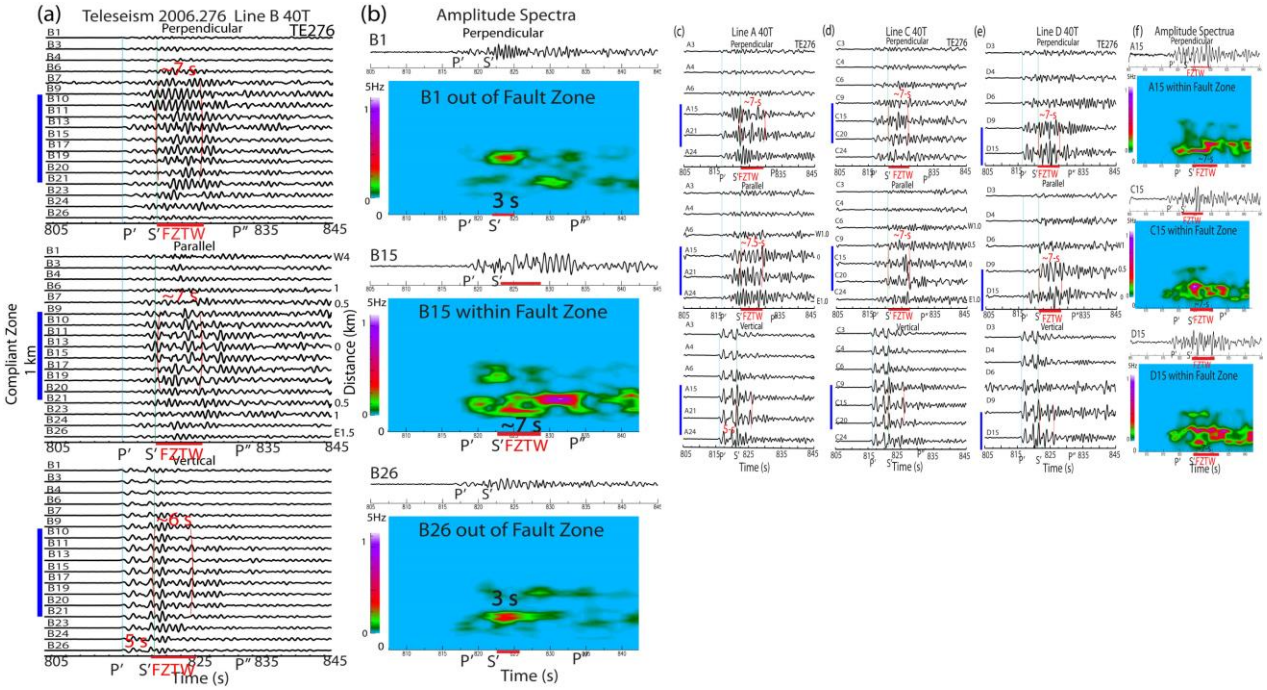


Fig.4 (a) Three-component seismograms recorded at seventeen 40T stations along Line B for a M6.3 teleseismic earthquake (TE276 in Table 1 and T276 in Fig. 1d). Teleseisms have been band-pass (0.5-2 Hz) filtered, and show prominent FZTWs with ~7-s post-S duration in two horizontal seismograms recorded at stations B9–B21 located within a ~1-km-wide CF compliant core zone. **(b)** Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs at 1-3 Hz with 6-s duration at on-fault station B15, but lower amplitude with shorter duration (~3-s) at stations B1 and B26 out of the compliant zone. **(c), (d)** and **(e)** Same as in (a), but at Lines A, C and D. Prominent FZTWs with ~7-s post-S duration in two horizontal seismograms recorded at stations located within the compliant zone. **(f)** Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs with ~7-s duration at on-fault stations A15, C15 and D15. We note multiple phases starting from P" in late coda of teleseisms.

Our observations indicate these wavetrains more likely to be the S-type FZTW arising from the S'-wave converted from P-wave at the Moho and then entering the bottom of the CF compliant zone at certain depth in the upper crust. We notice the P-type FZTWs immediately following the first P-arrivals

in vertical- and fault-perpendicular-component seismograms for this telemetry earthquake from which the ray path is nearly perpendicular to the fault strike. We notice multiple phases starting from P" and other phases in the late coda of teleseisms.

Then, we examine the data recorded at 40T stations along Line B for a M6.7 teleseismic earthquake (TE244 in Table 1) occurring at 38-km depth beneath Papua New Guinea Islands ~11,000 km (69.5°) west from the CF seismic array site in California (see Fig. 1d). P-waves arrive at ~848-s in the vertical-component seismograms. In teleseisms and spectral amplitude contours (Fig. 5a, b), we observe large-amplitude and long-duration (~7.5-s) wavetrains starting from ~5.5 s after the first-P arrivals at stations between B9 and B21 located within the ~1-km-wide CF compliant core zone but not at stations out of the zone, suggesting that these wavetrains are S-type FZTWs arising from the S'-wave converted from P-wave at the Moho and sub-vertically entering the bottom of the CF compliant zone at certain depth in the upper crust. These FZTWs are also observed on teleseisms recorded at stations of Lines A, B and C within the CF complaint zone (Figs. 5c, 6d). We notice the P-type FZTWs following the first P-arrival appearing in the vertical- and perpendicular-component seismograms from this shallow telemetry event from which waves propagate nearly perpendicular to the fault strike. We also notice multiple phases in the late coda on teleseisms. Fig. 5e shows receiver functions computed for teleseisms recorded at stations of Line B, Line A and Line C for this teleseismic earthquake. The peak amplitudes of P' and P-to-S converted wave S' from the Moho in receiver functions arrive at 0 s and ~5 s; the multiple-phase P" arrives at ~15 s. Large-amplitudes between 5 s and 12 s in receiver functions at stations B12, B13, A15 and C15) located within the CF compliant zone are likely related to S-type FZTWs arising from S' wave. In contrast, much lower amplitudes following S' in receiver functions at stations (B01, B26, A04 and C04) outside the compliant zone. We note that amplitude peaks following the P-arrivals registered at stations located within the CF compliant zone are likely P-type FZTWs.

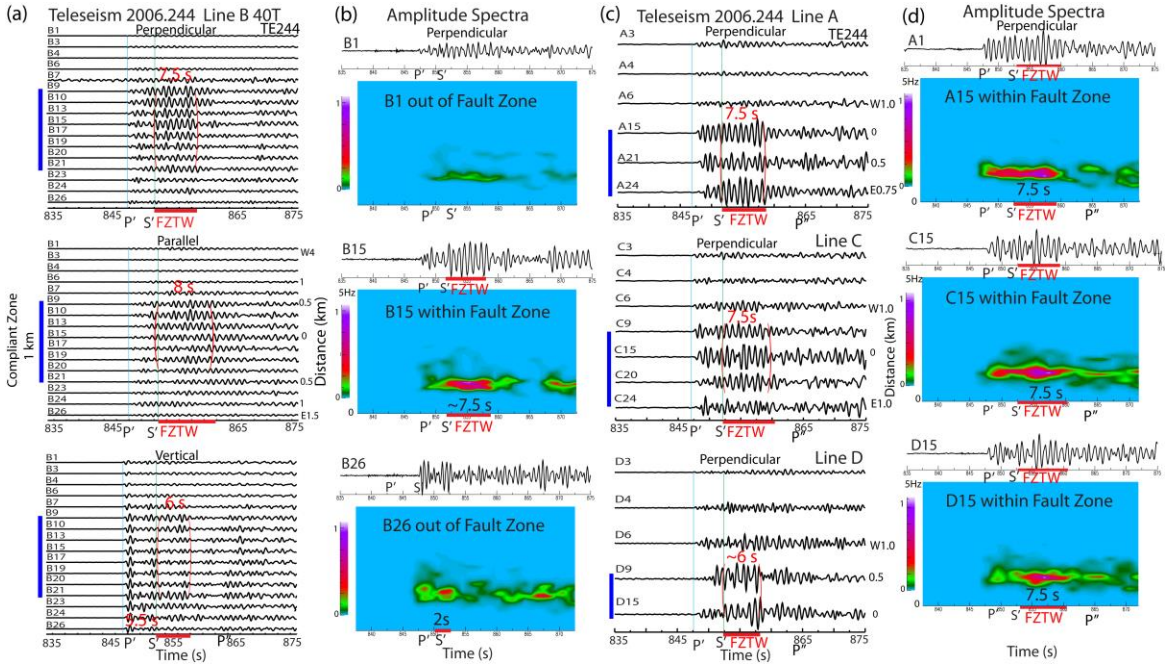


Fig.5 (a) and **(b)** Three-component teleseisms recorded at sixteen 40T stations along Line B for a M6.7 teleseismic earthquake occurring at 38-km depth (TE244 in Table 1 and Fig. 1d), showing prominent FZTWs with large amplitudes and ~7.5-s post-S duration starting at ~5.5-s after the first-P arrivals (in red brackets) on seismograms at stations B9–B21 located within a 1-km-wide compliant zone. Normalized spectral amplitude contours of teleseisms show large-amplitude (red color) FZTWs at 1-3 Hz with ~7.5-s duration at on-fault station B15, but lower amplitude with much shorter duration (2-s) at stations B1 and B26 out of the compliant zone. **(c)** Perpendicular-component teleseisms recorded at seismic Lines A, C and D. **(d)** Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs at 1-2 Hz with ~7.5-s duration at stations A15, C15 and D15 located on the CF surface trace.

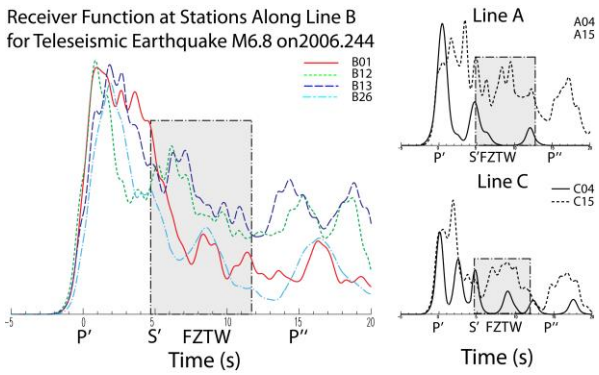


Fig.5 (e) Receiver functions computed by deconvolution between the radial- and vertical-component seismograms recorded at 4 stations of Line B and 2 stations of Line A and Line C for each. P'-arrivals are aligned at time 0 s. P-to-S converted waves (S') appear at ~5 s. The S-type FZTWs following S' show large amplitudes and ~7-s duration (in grey box) at stations (B-12, B13, A15 and C15) within the CF compliant zone while lower amplitudes at stations (B01, B26, A04 and C04) outside the compliant zone. The multiple P" arrives at ~15 s. Note that large amplitudes following P' and P" might relate to the P-type FZTWs.

Fig. 6 shows another example of this type of FZTWs recorded at both intermediate- and short-period stations along Line B for a M6.8 teleseismic earthquake (TE317 in Table 1 and Fig. 1d) occurring at 548-km depth in Bolivia, approximate 7900-km (78.9^{o}) southeast of the CF seismic array in California. Surface waves are invisible in teleseisms for this deepest earthquake in our study (see Fig.2). We observe FZTWs with ~5-6-s duration after the P-S converted S' waves on two horizontal-components of teleseisms recorded at stations within the ~1-km-wide CF compliant zone between B9 and B21 (Figs. 6a, b). FZTWs do not appear on the vertical-component seismograms because the ray path from this deep teleseismic earthquake to the CF array is nearly vertical. Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs at 1-3 Hz with ~6-s duration at on-fault station B15, but much weaker and shorter (2-s) post-S' wavetrains at stations B1 and B26 out of the compliant zone (Fig. 6c). We notice the P-type FZTW following the first P-arrival in teleseisms because waves from this deep teleseismic earthquake are nearly-vertical incident to the surface array. We notice that the durations (5-6 s) of FZTWs for this deep teleseismic event are shorter than those (7-8 s) for shallow events because the ray path from the deep event enters the bottom of the CF compliant zone more steeply than that from the shallow event, and thus FZTWs travel shorter distance within the compliant zone to arrive at the CF seismic array.

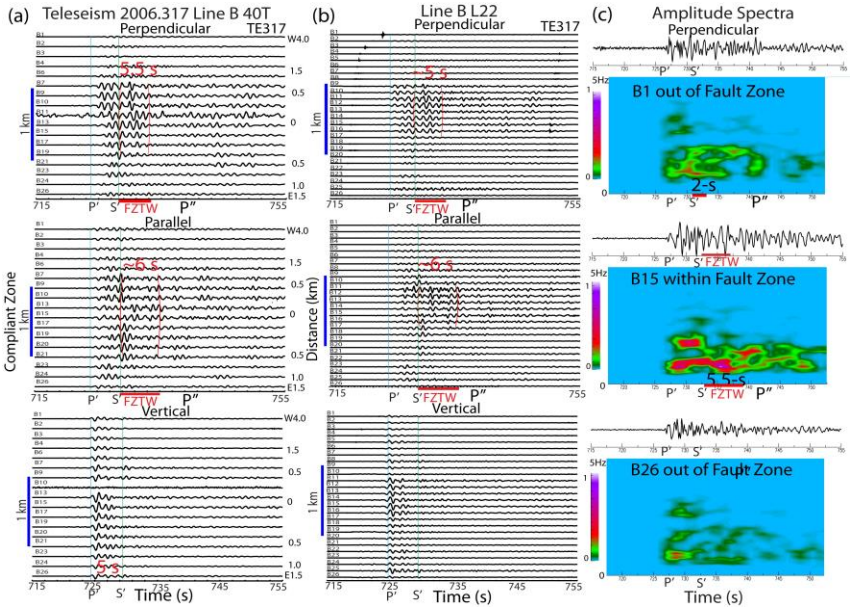


Fig.6 (a) and **(b)** Three-component seismograms recorded at sixteen 40T stations and twenty-six L22 2-Hz stations along Line B for a M6.8 teleseismic earthquake (TE317 in Table 1 and Fig. 1d). Teleseisms show prominent FZTWs with ~5-6-s post-S' duration in two horizontal seismograms recorded at stations B9–B21 located within a ~1-km-wide compliant zone. **(c)** Normalized spectral amplitude contours of teleseisms show large-amplitude FZTWs at 1-3 Hz with ~5.5-s duration at on-fault station B15, but lower amplitude with shorter duration (<2-s) at stations B1 and B26 out of the compliant zone.

Fig. 6d shows receiver functions computed for teleseisms recorded at stations of Line B, Line C and Line D for this teleseismic earthquake. The peak amplitudes of P' and P-to-S converted wave S' from the Moho in receiver functions arrive at 0 s and ~5 s, and the multiple-phase P" arrives at ~14-15 s. The large-amplitudes between ~5 s and 11.5 s in receiver functions at stations (B11, B13, B15, C15 and D15) located within the CF compliant zone are likely related to S-type FZTWs arising from S' wave. In contrast, much lower amplitudes following S' in receiver functions at stations (B03, B26, C03 and D03) outside the compliant zone. We note that amplitude peaks following the P-arrivals registered at stations located within the CF compliant zone are likely P-type FZTWs.

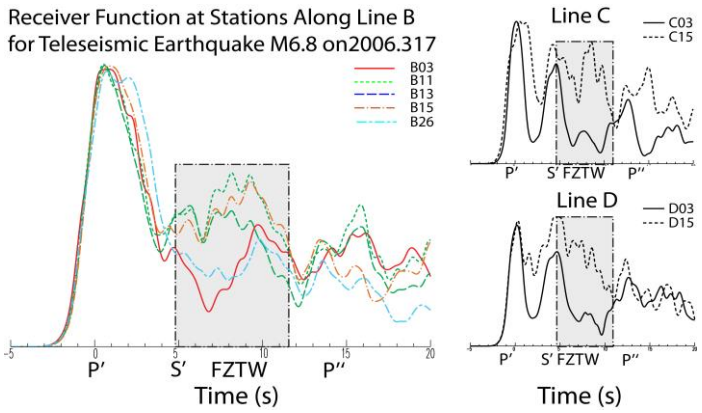


Fig.6 (d) Receiver functions computed by deconvolution between the radial- and vertical-component seismograms recorded at 5 stations of Line B and 2 stations of Line A and Line C for each. P'-arrivals are aligned at time 0 s. P-to-S converted waves (S') appear at ~4.5 s. The S-type FZTWs following S' show large amplitudes and ~6.5-s duration (in grey box) at stations (B11, B13, B15, C15 and D15) within the CF compliant zone while lower amplitudes at stations (B03, B26, C03 and D03) outside the compliant zone. The multiple P" arrives at ~15 s. Note that large amplitudes following P' and P" might relate to the P-type FZTWs.

Finally, we examine such type of FZTWs recorded at four seismic Lines A, B, C and D for two teleseismic earthquakes (TE272 and TE271 in Fig. 1d and Table 1) with magnitudes of M6.1 and M6.9, occurring at 53-km depth in South America, and 28-km depth beneath Tonga Islands in Pacific Ocean, ~6,200-km (55^{o}) southeast and ~8,700-km (73.5^{o}) southwest of the CF, respectively. Fig. 7A shows large-amplitude FZTWs with ~6-7-s post-S' duration in teleseisms recorded at stations with either 40T or L22 sensors located within the ~1-km-wide CF compliant zone, but not at farther stations out of the zone for teleseismic earthquake TE272. Fig. 7B shows large-amplitude FZTWs with ~7-8-s post-S' duration recorded at stations with either 40T or L22 sensors located within the ~1-km-wide CF compliant zone, but not at farther stations out of the zone for the shallow teleseismic earthquake TE271. The longer duration of FZTWs for this shallow event is probably because the ray path from the shallow event enters the bottom of the CF compliant zone less steeply than that from the deep event, and thus FZTWs travel longer distance within the compliant zone to arrive at the CF seismic array. We notice that P-type FZTWs following the first P-arrivals are obvious in parallel-component teleseisms for teleseismic event TE272 in Fig. 7A (c) because the ray path to the seismic array is sub-parallel to the CF strike. We also notice the multiples in the later coda, probably due to waves rebounded between the free surface and Moho discontinuity. We shall confirm this mechanism using receiver-function method.

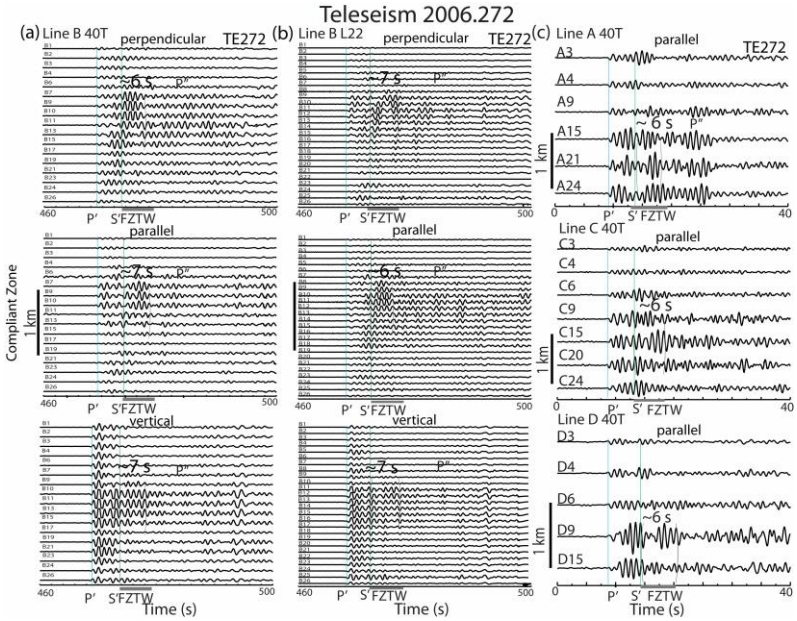


Fig.7A (a) and **(b)** Parallel-component seismograms recorded at seventeen 40T stations and twenty-six L22 stations along Line B for a M6.1 teleseismic earthquake (TE272 in Fig. 1d) occurring at 53-km depth. Teleseisms have been band-pass (0.5-2 Hz) filtered, showing prominent FZTWs with ~6-7-s post-S' duration in teleseisms recorded at stations located within the ~1-km-wide CF compliant zone. **(c)** Parallel-component teleseisms recorded at Lines A, C and D for this teleseismic event, showing prominent FZTWs with ~6-s post-S duration at stations within the compliant zone.

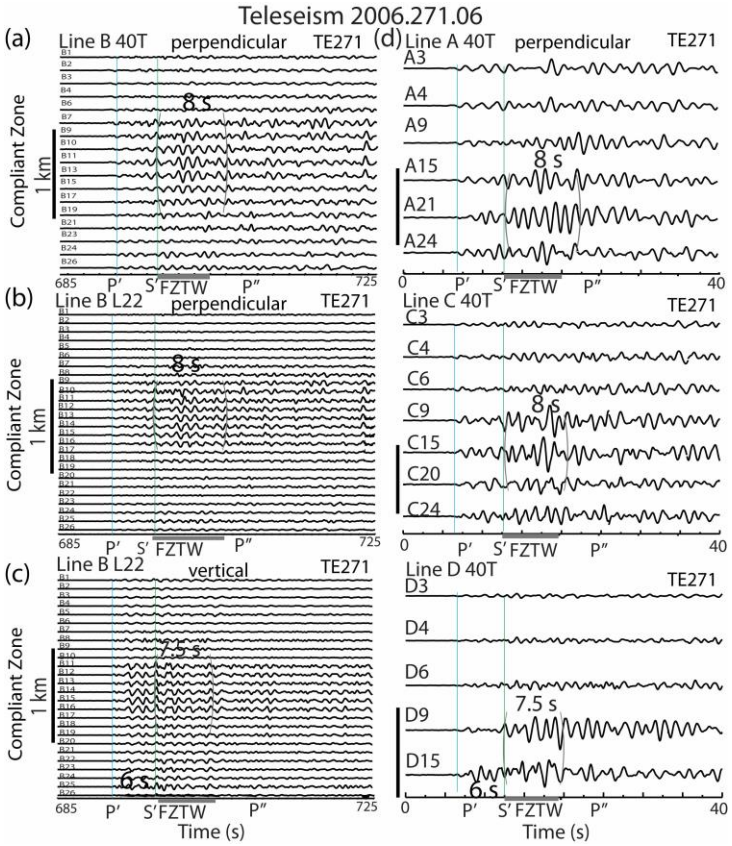


Fig.7B (a) Perpendicular-component seismograms at Line B with 40T sensors, **(b)** parallel-component and **(c)** vertical-component seismograms recorded at Line B with L22 sensors for a M6.9 teleseismic earthquake (TE271 in Fig. 1d) occurring at 28-km depth. Teleseisms have been band-pass (0.5-2 Hz) filtered, showing prominent FZTWs with ~8-s post-S' duration in teleseisms recorded at stations located within the ~1-km-wide compliant zone. **(d)** Perpendicular-component teleseisms recorded at Lines A, C and D, showing prominent FZTWs with ~8-s post-S duration at stations within the compliant zone.

In above-mentioned examples, seismograms recorded at the seismic array atop the Calico Fault show the consistent wavetrains with large amplitudes and approximate 5-8-s duration appearing in a time window starting from ~5-6 s after the P-wave first-arrival for teleseismic earthquakes (Fig. 3 to Fig. 7). We interpret these large-amplitude long-duration wavetrains arising from cohesive interference of seismic waves within the low-velocity waveguide along the CF compliant zone when seismic waves from teleseismic earthquakes enter the bottom of the compliant zone at certain depth. Since these wavetrains start at ~5-s after the first-arrival of P-wave (denoted by P' in figures), we tentatively interpret that the seismic waves entering the bottom of the CF low-velocity compliant zone are the P-S waves (denoted by S') converted at the Moho discontinuity at approximate 30-km depth (corresponding to the travel-time difference of ~5-s between P' and S'). Because these wavetrains are mostly apparent in horizontal-component teleseisms, they are likely the S-type of FZTWs with horizontal polarization. Their waveforms are similar to the fault-zone trapped waves generated by explosions and local earthquakes, and recorded at the same seismic array in the previous study [Cochran et al., 2009], but observed for teleseismic earthquakes.

3. 3-D Finite-Simulation of Observed FZTWs for Teleseismic Earthquake

We test a model of the Calico fault compliant zone from the surface to ~8-km depth on the basis of the previous model Cochran et al. (2009) combined with observations of FZTWs for teleseismic earthquakes. We use a 3-D finite-difference code (Graves, 1996; Vidale et al., 1985) to compute synthetic seismograms that fits teleseisms recorded at Line B across the CF. The finite-difference computer code is second order in time and fourth order in space, and it propagates the complete wave-field through an elastic media with a free surface boundary and spatially variable anelastic damping (an approximate Q). Our calculation used a 160-160-160 element grid in x–y–z coordinates, with a grid spacing of 62.5 m to simulate a volume of 10 km in width, 10 km in length, and 10 km in depth. The low-velocity waveguide, composed of a 1.5-km-wide fault zone of maximum velocity reduction of 50% sandwiched embedded in the higher-velocity surrounding rocks with a free surface. The receiver array was placed across the waveguide along the fault strike. In the first stage of our error-and trial forward modeling approach, we used 125-m grid spacing in a 10x10x10 km^3 volume to obtain coarse model parameters at low frequency to save computer memory space and computation time. We then used 62.5-m grid spacing to increase the resolution of fault-zone structure in the model at higher frequency. Fig. 8a shows the schematic diagram to illustrate the ray path of first-arrival P wave from a teleseismic event hitting the Moho discontinuity at 30-km depth, and transmitting P' and converted P-to-S (S') waves, which sub-vertically enter the bottom of low-velocity fault zone at 10-km depth to produce FZTWs within the fault zone. Fig. 8b exhibits finite-difference seismograms for a double couple source at the Moho with its epicenter 10-km with respect to the cross-fault array using the velocity model in Fig. 10c, in which the velocities within the fault-zone are reduced by 50% from wall-rock velocities. The low-velocity fault zone extends from the surface to the depth of ~8- km. Synthetic seismic waves are received at the array across the fault zone at surface, showing FZTWs characterized by large amplitudes and long durations arriving at stations located within a 1.5-km-wide low-velocity waveguide. Fig. 8b shows synthetic seismograms to compare with observed teleseisms recorded at Line B for a teleseismic earthquake TE272. The CF compliant zone with velocity reduction of 50% extends from the surface to 8-km depth. Synthetic waveforms developed from our inferred velocity model are in general comparable with observations for this teleseismic earthquake, suggesting that the depth extension of the CF compliant core zone with velocity reduction of 40-50% is approximately 8 km. The value is likely 2-3 km greater than the depth extension (5-6 km) of the Calico fault compliant core zone imaged by Cochran et

al. (2009). However, since the depth extension and the velocity reduction within the compliant zone are negotiable in the trial-and-error forward modeling, a systematic simulation of teleseisms recorded at the seismic array atop the CF for more teleseismic events with different incident angles to the CF seismic array will provide further constraints on the true depth extension of the CF compliant zone.

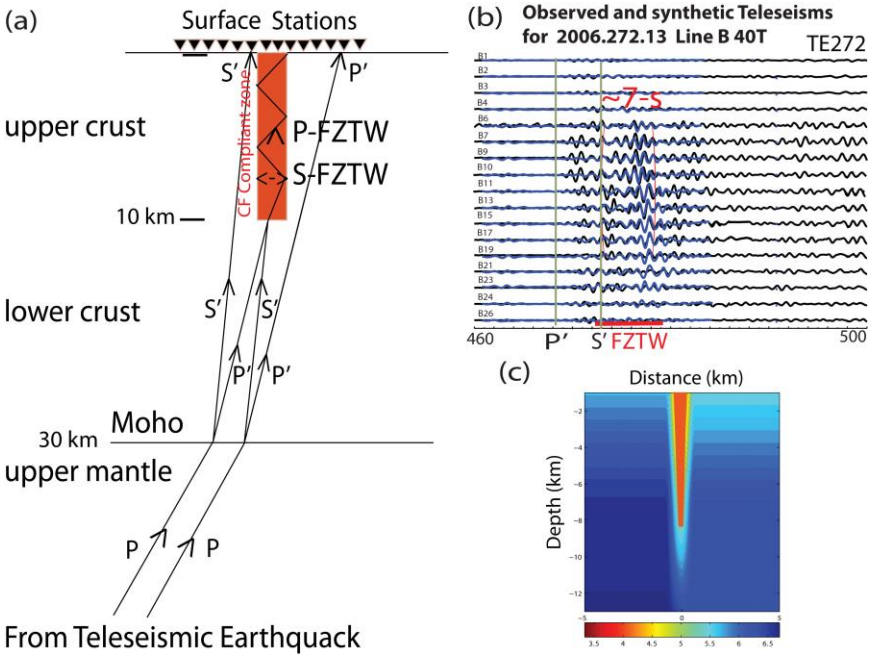


Fig. 8 (a) Schematic diagram illustrates the first-arrival P wave from teleseismic event hitting the Moho discontinuity between the lower crust and the upper mantle and transmitting P' and S' waves, which sub-vertically enter the bottom of low-velocity fault zone to produce P-type and S-type fault-zone trapped waves (P-FZTW and S-FZTW). They are recorded at stations (triangles) atop the CF compliant zone (marked by brawn color). **(b)** Observed (black lines) and synthetic (blue lines) perpendicular-component seismograms at Line B for a teleseismic earthquake TE272 in Table 1. The CF compliant zone with velocity reduction of 50% extends from the surface to 8-km depth. A double-couple source is put on the Moho at 30-km depth and with 10-km epicentral distance from the receiver array. **(c)** The model of P-wave velocities across the Calico fault with a low-velocity waveguide along the CF compliant zone to ~8-km depth, within which velocities are reduced to the maximum of ~40-50% in the center of a ~1-km-wide fault core zone at shallow depth (Cochran, et al., 2009). The fault zone is ~1.5 km wide at the surface. The lateral velocity profile across the fault is approximated as a Hanning taper, and the velocity reduction tapers linearly to zero between 0 and 10 km depth.

5. Discussion

An outstanding question is the depth extent of the low-velocity damage zone along the fault. Some researchers argue that the low-velocity damage zone on faults is a near-surface feature that reaches only down to the top of seismogenic zone at the depth less than 2-3 km (e.g. *Ben-Zion et al*. 2003; *Lewis et al*. 2010). Others argue that it extends across seismogenic zone at depths to ~10-km depth (e.g. *Korneev et al*. 2003; *Li et al.*, 2000; *Li and Malin* 2008). *Wu et al*. (2010) showed that the low-velocity waveguide on the SAF at Parkfield extends to the depth of 10 km or more using SAFOD borehole data. Recently, *Ellsworth and Malin* (2011) document a profound zone of rock damage on the Parkfield San Andreas fault downwards to at least half way (>5-6 km) through the seismogenic crust using both P-type and S-type of FZTWs recorded at the SAFOD mainhole seismograph. FZTWs recorded at Parkfield surface and borehole stations shows that the LVZ on the Parkfield SAF extends to the depth of at least ~7-8 km although the velocity reduction within the damage zone decreases with depth due to the increasing confining pressures. Although FZTWs generated by explosions and local earthquakes have been used for characterization of the subsurface fault damage structure, the wave propagation coverage is lake of resolution for the deep fault zone structure because these seismic sources are located at shallow depths.

In this article, we introduce a new-type of FZTWs recorded at the dense seismic array atop the Calico fault zone for teleseismic earthquakes. We interpret the observed FZTWs arise from S-wave converted from P-wave at the Moho discontinuity and sub-vertically incident at the bottom of CF compliant zone. Therefore, these waves are capable to provide more unprecedented constraints on the depth extension of fault damage structure at seismogenic depth where there is lack of approach of waves generated by explosions and local earthquakes.

The data recorded at the dense array of 100 three-component seismographs deployed atop the Calico fault in the central Mojave Desert, Southern California, provide us a good opportunity to detect the seismic phases closely following the first-arrival P wave. The seismic array worked in continuous recording mode at an extremely quiet site for six months and recorded 72 teleseismic earthquakes with magnitudes $M \geq 6$ and occurring at distances of ~40^o-80^o great circles (approximate 5,000-11,000 km) to the Calico seismic array site. In order to further confirm our observation of this new type of FZTWs, We shall simulate this new type of FZTWs using a full waveform modeling technique with inverse, such as the FD3D tomography method (Chen, 2012). We shall also identify the FZTWs from teleseismic earthquakes recorded at seismic arrays deployed atop other active faults such as San Andreas Fault near Parkfield (Li et al., 2004; 2012) and at the rupture zone of the 2010 M7.1 and 2011 M6.3 Canterbury earthquakes in New Zealand (Li et al., 2014).

Acknowledgments

This study is supported by Southern California Earthquake Center's award 15195 and is benefited by the previous results by Cochran et al. (2009). Thank Elizabeth Cochran for her work on archive of the data recorded at the seismic array deployed at the Calico fault in 2006. Thank En-Jui Lee for his help to download the data from IRIS DMC and Y.L. Chen for using his computer program to calculate receiver functions.

Reference

Ben-Zion, Y., Z. Peng, D. Okaya, L. Seeber, J. G. Armbruster, N. Ozer, A. J. Michael, S. Barris, and M. Aktar, A shallow fault zone structure illuminated by trapped waves in the Karadere-Dusce branch of the North Anatolian Fault, Western Turkey, Geophys. J. Int., 152, 699-717, 2003.

Chen, P., Full-wave seismic data assimilation: a unified methodology for seismic waveform inversion, in Imaging, Modeling and Assimilation in Seismology, edited by Yong-Gang Li, 350pp, published by DE GRUYER and China Higher Education Press, Feburary 2012. ISBN 3110259028, 9783110259025.

Cochran, E. S., Y. G Li, P. M. Shearer, S. Barbot, Y. Fialko, and J. E. Vidale, Seismic and geodetic evidence for extensive, long-lived fault damage zones, Geology, v9.37; no. 4, 315–318, doi:10.1130/G25306A.1; Data Repository item 2009082, 2009.

Ellsworth W. L. and P. E. Malin, Deep rock damage in the San Andreas fault revealed by P- and S-type fault zone guided waves, Sibson's volume, Geological Society Special Publication, New Zealand, V359, 39-53, 2011.

Fialko, Y., Probing the mechanical properties of seismically active crust with space geodesy: Study of the coseismic deformation due to the 1992 Mw7.3 Landers (southern California) earthquake: J. Geophys. Res., v. 109, B03307, doi: 10.1029/2003JB002756, 2004.

Fialko, Y., Sandwell, D., Agnew, D., Simons, M., Shearer, P., and Minster, B., Deformation on nearby faults induced by the 1999 Hector Mine earthquake: Science, v. 297, p. 1858–1862, doi: 10.1126/science.1074671, 2002.

Ganev, P.N., Dolan, J.F., Oskin, M.E., Owen, L.A., and Le, K.N., Paleoseismologic evidence for multiple Holocene earthquakes on the Calico fault: Implications for earthquake clustering in the eastern California shear zone: Southern California Earthquake Center Annual Meeting Proceedings and Abstracts, v. XVIII, p. 153–154, 2008.

Graves, R.W., Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, Bull. Seis. Soc. Am., 86, 1091-1106, 1996.

Korneev, V. A., R. M. Nadeau, and T. V. McEvilly, Seismological studies at Parkfield IX: Fault-zone imaging using guided wave attenuation, Bull. Seism. Soc. Am. 80, 1245-1271, 2003.

Lewis, M. A. and Y. Ben-Zion, Diversity of fault zone damage and trapping structures in the Parkfield section of the San Andreas Fault from comprehensive analysis of near fault seismograms. Geophys. J. Int., doi: 10.1111/j.1365-246X.2010.04816.x. 2010.

Li, Y.G., P.C. Leary, K. Aki, and P.E. Malin, Seismic trapped modes in the Oroville and San Andreas fault zones, Science 249, 763–766, 1990.

Li, Y.G., and P.C. Leary, Fault-zone trapped seismic waves, Bull. Seis. Soc. Am., 80, 1245-1271, 1990.

Li, Y.G. and J.E. Vidale, Low-velocity fault zone guided waves; numerical investigations of trapping efficiency, Bull. Seis. Soc. Am., 86, 371-378, 1996.

Li，Y. G. and P. E. Malin, San Andreas Fault damage at SAFOD viewed with fault-guided waves, Geophys. Res. Lett., 35, L08304, doi:10.1029/2007GL032924, 2008.

Li, Y. G., K. Aki, D. Adams, A. Hasemi, and W. H. K. Lee, Seismic guided waves trapped in the fault zone of the Landers, California, earthquake of 1992, J. Geophys. Res., 99, 11,705-11,722, 1994.

Li, Y. G., J. E. Vidale, K. Aki, and F. Xu, Depth-dependent structure of the Landers fault zone using fault zone trapped waves generated by aftershocks, J. Geophys. Res. 105, 6237–6254, 2000.

Li, Y. G., J. E. Vidale, S. M. Day and D. Oglesby, Study of the M7.1 Hector Mine, California, earthquake fault plan by fault-zone trapped waves, Hector Mine Earthquake Special Issue, Bull. Seism. Soc. Am., 92, 1318-1332, 2002.

Li, Y. G., J. E. Vidale, D. Oglesby, S. M. Day and E. Cochran, Multiple-fault rupture of the M7.1 Hector Mine, California, earthquake from fault-zone trapped waves, J. Geophys. Res., 108, ESE 11, 1-25, 2003.

Li, Y. G., J. E., Vidale, and S. E. Cochran, Low-velocity damaged structure of the San Andreas fault at Parkfield from fault-zone trapped waves, Geophy. Res. Lett., 31, L12S06, 2004.

Li, Y. G., P. Malin, and E. Cochran, Fault-zone trapped waves: High-Resolution Characterization of the Damage Zone on the Parkfield San Andreas Fault at Depth, Chapter 3 pp 107-150 in book: Imaging, Modeling and Assimilation in Seismology, edited by Y. G. Li, China High-Education Press, Beijing, De Gruyter, Boston, February 2012.

Li, Y.G., G. De Pascale, M. Quigley, and D. Gravely, Fault damage zones of the M7.1 Darfield and M6.3 Christchurch earthquakes characterized by fault-zone trapped waves, Tectonophysics, V618, 79-101, (10.1016/j.tecto.2014.01.029), 2014.

Peltzer, G., Crampe, F., Hensley, S., and Rosen, P., Transient strain accumulation and fault interaction in the Eastern California shear zone: Geology, V29, 975–978, doi: 10.1130/0091–7613(2001)029<0975: TSAAFI>2.0.CO;2, 2001.

Vidale, J.E., D.V. Helmberger, and R.W. Clayton, Finite-difference seismograms for SH waves, Bull. Seis. Soc. Am. 75, 1765-1782, 1985.

Wu, J., J. A. Hole, and J. A. Snoke, Fault-zone structure at depth from differential dispersion of seismic guided waves: Evidence for a deep waveguide on the San Andreas fault, Geophys. J. Int., 182, 343-354, doi: 10.111.j.1365-246X.2010.04612.x. 2010.

ASSESSING STUDENT LRARNING IN ARITHMETIC AND NUMBER SYSTEMS

Shuhua An

California State University, Long Beach, USA

This paper analyzes trends in assessing student learning arithmetic and number systems by discussing two different types of assessments, followed by addressing new trends in assessments in arithmetic and number systems. The paper reviewed and highlighed examples of studies in assessment in conceptual undersanding, procedual fluency, reasoning, and problem sovling, and the paper also reviewed examples of assesment in combined mulitple areas and classroom assessment.

Introduction

One of the key instructional components for effective elementary mathematics instruction is assessment. National Research Council (1993) specifies mathematics assessment as "a way of measuring what students know and of expressing what students should learn" (p.1). Assessment is not only "a fact-finding activity" that describes conditions that exist at a particular time (Best & Kahn, 2006), but also should support the learning of important mathematics and furnish useful information to both teachers and students (NCTM, 1995). This important role of assessment requires educators to consider assesment as "the process of gathering evidence about a student's knowledge of, ability to use, and disposition towards mathematics and of making inferences from the evidence for a variety of purposes" (NCTM, 1995, p.3). The purpose of assessment as a process of gathering evidence and making inferences from that evidence for a variety of goals has played a vital role in mathematics assessment in the last two decades. Recent developments in the new Common Core State Standards for Mathematics (CCSSM) the US have futher heightened the clear purpose of assessment as to support and enhance student learning and called for focusing on mathematics instruction that provide. The three the key shifts are called for by the Common Core: (1) **Foucs: Greater focus on fewer topics, (2) Coherence: Linking topics and thinking across grades, and (3) Rigor: Pursue conceptual understanding, procedural skills and fluency, and application with equal intensity (**National Governors Association Center for Best Practices, & Council of Chief State School Officers, 2010). The key shift of rigor provided a clear direction in mathematics instruction as well as assessment that requires students to achieve a balance in mathematics learning in conceptual understanding, procedural skill and fluency, and the ability to apply mathematics to solve problems (Wu, 2008; An & Wu, 2014), which is supported by research that effective classroom teachers always use multiple forms of assessments that are meaningful and applicable in classroom to support student learning (An & Wu, 2014; McMillan, 2011).

The purpose of this article is to identify and describe examples of the central themes in order to illustrate key aspects of the assessments in arithmetic and number systems. This article provides the overall review on traditional and alternative assessments and addresses new trends and chellenges in assessments in arithmetic and number systems, followed by discussing some specific assesment studies based on a review of related literature on assessing student matheamtical conceptual understanding, procedural skill and fluency, and problem solving teacher in arithmetic and number systems in the recent decades. The article also show an examples of combination of the three areas, and discussion of the classroom assessment.

Traditional and alternative assessments in arithmetic and number systems

Two main types of assessments have been used in assessing student learning in arithmetic and number systems, traditional and alternative. Traditional assessments, such as standardized tests are mainly multiple-choice tests. According to Thornberg (2001), standardized testing has become a large part of shaping the educational system and its improvements. They measure test-taking skills and a certain degree of what students know, and they cover knowledge-based questions, which students can answer by memorization of factual information, or comprehension, simple problem solving, or even good guessing

(Fuchs et al. 1999). However, because of their accountability, educators often feel pressured to teach to the test with the hopes that students will score high (Lewis, 2001). The results of standardized tests, such as NAEP mathematics from 1990–2013 show an improving trend of elementary childrens' learning ourcome, but still no more than 42% percentage of fourth-graders were at or above *Proficient level (The Nation's Report Card, 2013).* Researchers (Suurtamm, Koch, & Arden 2010) have called that assessment must go beyond focusing on memorization of algorithm and procedures by responding and assessing students' mathematical understanding. Therefore, standardized testing should not be the only form of assessment used in schools (NCTM, 2000).

In recent decades, the other types of assessments - alternative assessment is becoming an increasingly important area in mathematics education because they are authentic assessments, performance-based (Sweet, 1993), or formative assessments, providing a more reflective evaluation of what students are learning in the classroom, as opposed to a general overview which standardized testing measures. According to Wu and An (2007), alternative assessments offer a wider range of student choice, and allow teachers to monitor their students' progress in a much more detailed manner. They offer a more student-centered approach to learning, and measure students' ongoing progress in multiple forms. For example, mathematics portfolios of students' work throughout the year are used to mark growth and improvement, while mathematics journals promote students to use their thinking skills and write about their learning (Micklo, 1997).

One of the most significant discussions in assessments, no matter what forms of assessments, whether traditional or alternative, concerns which characteristics of assessments work for the students and which characteristics hurt the students. After all, students are at the center of education and their progress marks the quality of education (Wu & An, 2007). In order to appropriately assess student learning in arithmetic and number systems, it is important to know what the assessment entails and what the assessments are actually measuring.

New trends and chellenges in assessments in arithmetic and number systems

Smarter balanced assessment and four claims. Recent developments and impletementations in the common core standards for mathematics (CCSSM) in the United States have led to a renewed interest in assessing students' higher-order thinking with deep understanding and reasoning in mathematics in various froms. For examples, the new Smarter Balanced Assessments, alinged CCSSM, require students to think critically, solve problems, and show a greater depth of knowledge (Smarter Balanced Assessment Consortium, 2012). The Smarter Balanced Assessment has four claims: (1) Concepts and Procedures: Students can explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency. (2) Problem Solving: Students can solve a range of complex, well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. (3) Communicating Reasoning: Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. (4) Modeling and Data Analysis: Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems (The Smarter Balanced Assessment Consortium, 2012) The unique feature of the Smarter Balanced summative assessments is that it goes beyond multiple-choice questions to include four types of items: 1) Selected-response items, 2) Technology-enhanced items, 3) Constructed-response items, and 4) Performance tasks. However, It is a chellenging task for elementary teachers to support deeper learning and the development of 21st century competencies essential for students' future success (Pellegrino & Hilton, 2012). Drawing on Norman Webb's depth of knowledge (DOK) (Webb, 1997), Herman and Linn (2013) define deep learning in four DOK levels: DOK1- Recall and Reproduction; DOK2 - Basic Application of Skills/Concepts; DOK3 - Strategic Thinking; DOK4 - Extended Thinking. These four levels of DOK now is widely used by classroom teachers in assessing and supporting deep learning in mathematics for their students in the United States.

Assessing four claims based on CCSSMPs. The new Common Core State Standards for Mathematical Practice (CCSSMPs) in the US describe a variety of expertise that mathematics educators at all levels

should seek to develop in their students **(**National Governors Association Center for Best Practices, & Council of Chief State School Officers, 2010). These CCSSMPs require the development of deep understanding and reasoning in mathematics with a high level of mathematics thinking, reasoning, arguments, communication, and modeling skills. However, these CCSSMPs have posed challenges to classroom teachers on how to forster students' deep understanding, sound reasoning and argument skills in math classrooms. In recent released 2015 Smarter Balanced Assement results in California show that more than 40% of elementary students in grades 3-5 scored below standards in the areas of concepts and procedures, followed by more than 38% below standards in problem solving **/Modeling and Data Analysis. The area of Communicating Reasoning had more than 39% of students who did not met standards at grades 4 and 5.**

Areas	Below Standard	3rd	4th	5th
Concepts and Procedures	**The student does not demonstrate the ability** to explain and apply mathematical concepts and interpret and carry out mathematical procedures with precision and fluency.	40 %	47 %	51 %
Problem Solving/Modeling and Data Analysis	**The student does not demonstrate the ability** to solve a range of complex, well-posed problems in pure and applied mathematics, making productive use of knowledge and problem-solving strategies. **The student does not demonstrate the ability** to analyze complex, real-world scenarios and construct and use mathematical models to interpret and solve problems.	38 %	39 %	46 %
Communicating Reasoning	**The student does not demonstrate the ability to** clearly and precisely construct viable arguments to support the student's own reasoning and to critique the reasoning of others.	28 %	39 %	41 %

Table 1: California Elementary Student Smarter Balanced Assessment Results

These data show an urgent need to develop an effective approach to support elementary children in improving their learning in arithmetic and number systems as they are the major standards in CCSSM in the US.

Performance task and Smarter balanced assessment

Another challenge faced by many classroom teachers is supporting students in communicating reasoning and doing performance tasks bescasue Smarter Balanced Assesements include performance task items in its tests at every grade level. Performance- assessments are alternative or authentic assessments (Sweet, 1993). The aim of the authentic assessments is determine what students have learned, and how *they are to apply* their knowledge to a variety of tasks (Adamson & Darling Hammond, 2010; Caffrey, 2009). According to the Smarter Balanced Assessment Consortium (2012), the p**erformance tasks of** Smarter Balanced Assesements measure a student's ability to demonstrate critical-thinking and problem-solving skills. They challenge students to apply their knowledge and skills to respond to complex real-world problems. The forms of the p**erformance tasks** can be best described as collections of questions and activities that are coherently connected to a single theme or scenario. Therefore the goals of performance tasks are to measure students' capacities of deep learning, understanding, research and writing skills, critical thinking and reasoing, and complex analysis. To support students' learning and success in performance tasks, students should be given meaningful and real world related tasks with authentic products to show their understanding of the material in performance tasks (McBee & Barnes, 1998). It provides teachers with the opportunity to promote students' interests and engage them in active learning through reflection and demonstration of the thinking processes (Tung, 2010). These student-centered assessments evaluate students at a direct level (Wright, 2001).

Examples of assessments in conceptual understanding, compuation, and problem solving

This section provides examples of asessments focusing conceptul understanding, procedual fluencey, and problem solving, and their impact on teaching and learning mathematics in arithmetic and number systems. The samples of studies were selected and catergorized based on themes and internatioal representation to show trends in assessments in the specific foucses. These samples are orgnized based on five themes: understanding, procedual fluencey, problem solving, combination of the three areas, and classroom assessment.

Assessment in conceptual understanding arithmetic and number systems

Berman's study (2011) aimed to help teachers assess students' understanding of place value, so teachers can ensure what conceptual learning and teaching in arithmetic are needed. The study described how to assess students' understanding and manipulation of number system by using the sixTasksofPlaceValue (SToPV), which takes about five minutes to do, and can be used to explore students from third grade. The tasks were a combination of doing some hands-on activities and giving responses to the teacher's questions. There were three stages of conceptual understanding in place value in the assessment, understanding, construction, and emergent. Each task gave the rubric of the four different levels of conceptual understanding. According to student's response levels in each task, teachers could determine which stage the student was demonstrating by referring to the definition of the three stages. The author conclused that the SToPV can be administered in daily classroom, and it gives clear definition of where the student is in conceptual understanding of the number system, which helps classroom teachers in their math teaching.

The study of Baglici, Codding, and Tryon (2010) followed a cohort of the first grade students for a whole school year, which was an extension of a previous study of the Tests of Early Numeracy (TEN; Chard et al., 2005; Clarke & Shinn, 2002, 2004). The purpose of the study was to determine which TEN measures demonstrate sensitivity to growth from kindergarten to first grade, ascertain whether TEN performance in kindergarten is correlated with first-grade TEN performance, and examine whether TEN performance in kindergarten predicts first-grade mathematics computation skills, year-end mathematics report card grades, and teacher ratings of mathematics skills. The participants were 61 students from three suburban public schools near New York City. The study assessed the student performance from kindergarten to first grade. The assessment used four 1- minute TEN measures to collect students' response in arithmetic. The four measures were oral counting, number identification, quantity discrimination, and missing number. Responses for the four measures ranged from 0 to 10 for kindergarten and up to 20 for first grade. Findings showed that "In nearly all the analyses, performance on the kindergarten MN measure was a significant predictor of first-grade computation and teacher-determined outcome measures (i.e., ACES-M); it also yielded accept- able reliability coefficients" (p. 100).

Bailey, Hoard, Nugent, and Geary's study (2012) was designed to determine "whether measures of conceptual knowledge of fractions and computational fractions skills concurrently predict mathematics achievement, controlling for other factors" (p. 448) by providing a cross-lagged assessment to connect fractions and mathematics achievement across sixth and seventh grades. The study analyzed 212 children who completed at least two tasks given by the researchers. The standardized measures consist of intelligence, achievement, working memory, and mathematical tasks (number sets and number line estimation). The psychometric measures consisted of computational arithmetic, computational fractions, and fractions comparison test. The cross-lagged effects showed that performance on the sixth grade fractions concepts measure predicted 1-year gains in mathematics achievement ($ß = .14$, $p < .01$), controlling for the central executive component of working memory and intelligence, but sixth grade mathematics achievement does not predict gains on the fractions concepts measure ($ß = .03$, $p > .50$). The research also demonstrated that measures of fluency with computational fractions significantly predicted seventh grade mathematics achievement above and beyond the influence of fluency in computational whole number arithmetic, performance on number fluency and number line tasks, central executive span, and intelligence. The findings indicated the importance of fraction competence for mathematics achievement beyond the elementary school years (Siegler et al., 2011, 2012) and supported the hypothesis

that improving children's competence with fractions is likely to facilitate gains in mathematics achievement.

Assessment in computation and procedural fluency in arithmetic and number systems

Liu's study (2009) aimed at collecting baseline data for computational estimation by Chinese students. About 403 third graders and fifth graders participted in the study from four different types of elementary schools in different areas in China. The instrument of the study was based on the Multiplication Estimation Test (MET) with some revisions to the original one. The revised MET had 32 items with an equal number of problems on all the dimensions. The author concluded that students often will not estimate simply at the request to estimate if an exact answer is within their mental computation capability, and a two-step process is suggested for helping students decide what route to take when given arithmetic problems.

The longitudinal study by Muldoon, Towse, Simms, Perra, and Menzies (2013) aimed at assessing the relationships between number line estimation, counting, and mathematical abilities, in response to claims that the quality (and in particular linearity) of children's mental representation of number acts as a constraint on number development. The pariticipants were 99 five-year-old students from four primary schools in Edinburg, Scotland. They were tested individually using the three tasks in number estimation, counting, and math ability in four testing sessions with varied focuses in each session 3 monthly intervals. For all analyses the researchers used the ability score. The finding showed that there were correlations between the three types of ability, but while the quality of children's estimations changed over time and performance on the mathematical tasks improved over the same period, changes in one were not associated with changes in the other. The findings suggested that the linearity of number representation is not significantly privileged in its impact over and above simple procedural number skills. The authors proposed that both early arithmetic success and estimating skill are bound closely to developments in counting ability.

Assessment in problem solving in arithmetic and number systems

Kalyuga (2006) described an alternative schema-based rapid assessment technique and investigated how to apply it in arithmetic word problem solving in the study. Fifity five eighth graders from a Sydney school participatec in the assessment. The students were asked to finish two different tests, the traditional test and the first-step diagnosis test. In the first-step diagnosis test, 20 arithmetic problems were given, and students only needed to write their first solution in 30s. If a response to a task was an immediate next step expected in the fully worked out, detailed fine-grained solution procedure according to the first schema for that task, the response was allocated a score 1. If an answer was not an immediate next but one of the following steps towards the solution (or even the final step of the solution), it was al- located an additional score for each skipped step. The study showed that a rapid measure of learners' schematic knowledge structures based on monitoring immediate traces of cognitive processes in working memory can be constructed and used for valid diagnostic inferences in a relatively complex instructional domain.

A case study by Mundia (2012) aimed at identifying the degree and nature of problems in math for fifth graders, finding and recommending ways the child's math problems could be investigated further to gain additional insights. The participants were 29 Grade 4 children. The data for this case study were collected through observations, school assessment reports and documents, an in-take interview with one of the parents, a researcher constructed 16- item diagnostic test covering contents (addition, subtraction, multiplication, division); an error analysis of the child's mistakes in mathematics from the diagnostic class wide test; and a think-aloud diagnostic interview based on the error analyses. The survey data were analyzed quantitatively while observational and interview data were analyzed qualitatively. The study identified some learning difficulties of the fourth graders. The difficulties were "inability to use the four arithmetic operations (addition, subtraction, multiplication, and division) efficiently; not understanding the relationship between units, tens and hundreds; using any two of the four arithmetic processes (+, - , x,

÷) in combination within one operation; treating each column as a separate problem; place value problems or wrong alignment of numbers; poor eye-hand coordination leading to dysgraphia; and short-term memory / memory lapses" (p. 361).

Assessing student learning in muliple areas and forms in arithmetic and number systems

The effective assessment task should be an open-ended task, and it should include the three components (NRC, 1993): 1) Involving significant mathematics; 2) Enhancing mathematics learning and supporting good instructional practice; 3) Supporting every student's opportunity to learn important mathematics. Various studies suggest using rich mathematical tasks to develop students' capacity in reasoning and argument (Mok & Kaur, 2006; Shimizu, Kaur, Huang, & Clarke, 2010). Mathematics tasks are important vehicles for classroom instruction that aims to enhance students' learning. To achieve quality mathematics instruction, then, the role of mathematical tasks to stimulate students' cognitive processes is crucial (Hiebert & Wearne, 1993). Current research evidence indicates that students who are given opportunities to work on their problem solving skills in rich mathematical tasks enjoy the subject more, are more confident and are more likely to continue studying mathematics, or mathematics-related subjects (Hewson, 2015). One example of rich mathematical tasks in arithmetic and number systems demonstrated in Wu and An's study (2015) is using the Model-Strategy-Application with Reasoning (MSAR) in assessing the effects of children learning in diverse classrooms. The results shows that applying the MSAR approach in teaching and assessing student learning helped diverse students at grades 2, 3, and 5 understand arithmetic and number systems from drawing visual models, developing procedural fluency using various strategies, and building competence in problem solving in real world applications. The findings of Wu and An's support Hopkins's (2011) idea of combining concepts and procedures and problem solving together. Hopkins (2011) indicated that students will have mathematical learning difficulties when some students' conceptual understanding and procedural learning of arithmetic becomes less integrated. For example, various research studies showed the effects of integral nature of understanding place value numeration with applying it in arithmetic and problem solving on children's learning place value acknowledged (Steffe & Cobb, 1988; Wright, 1996; Hopkins; 2011).

Seethaler and Fuchs' study (2006) aimed at examining the relations of various cognitive abilities and aspects of math performance with computational estimation skill of third graders. Students (n = 315) were assessed on language, nonverbal reasoning, concept formation, processing speed, long-term memory, working memory, inattentive behavior, basic reading skill, arithmetic number combination skill, double-digit computation skill, and computational estimation ability. The researchers conducted a one-way analysis of variance (ANOVA) on estimation skill, using computation performance status as the between-groups factor. Then, correlation and multiple regression were used to explore the extent to which the predictor variables under examination correlated with, and accounted for variance in, estimation skill. The conclusion of the study was that skills in arithmetic number combinations, nonverbal reasoning, concept formation, working memory, and inattentive behavior may play significant roles in the prediction of estimation outcome.

Classroom Assessment

According to Callingham (2010), "The classroom is the powerhouse of learning" (p.2). Teachers play a major role in making student learning successful (Hattie, 2009). Sandra Horn, and William Sanders (1997) have corroborated the conclusions and indicated that the most important factor affecting student learning is the teacher. It is important to provide a varity of assesments in classroom to support student mathematics learning. These type of assessments include preassessent that address a student readiness to learn to support planning for achieving maximally effectiveness in student learning, and formative assessment that occurrs during teaching (McMillan, 2011). Formative assessment is an interactive and ongoing process (Airasian, 1994) and focuses on active feedback by teachers to support learning (Marzano, 2006; Shavelson, 2006). According to Marzano (2006), research on effective classroom assessment lead to the following four gereralizations:

- Feedback from classroom assessments should give students a clear picture of their progress and how they might improve.
- Feedback on classroom assessments should encourage students to improve their learning.
- Classroom assessment should be formative in nature.
- Formative classroom assessments should be frequent.

A study on formative classroom assessment conducted by An and Wu (2011) showed that teachers can provide more and better feedback to students when they use the sampling strategy addressed in the study for assessing and analyzing student work. Their study engaged ten classroom teachers in an inquiry process of the four steps of identifying errors, analyzing reasons for the errors, designing approaches for correction, and taking action for correction. The teachers, include a 5th grade teacher, made obvious progress in their knowledge of assessing students' thinking, understanding the difficulties and challenges their students had in learning mathematics through analyzing their daily homework.

Summary and reflection

In summary, the list of studies reviewed in this artilce is not intended to be exhaustive. The purpose of this article was to address and highlight some examples of studies related to assessment in student learning in arithmetic and number systems.

The results of these studies demenstrated recent trends in assesments in arithmetic and number systems, focusing the process of assessing student learning in conceptual undersanding, procedual fluency, reasoning, and problem sovling in real world application with mulitple forms. These studies showed that classroom teachers are the key in assessing and supporting studnet learning, and effective classroom teachers should use multiple forms of assessments that are interactive and meaningful in classroom to support student learning in arithmetic and number systems.

.

References

Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw Hill

An. S. & Wu, Z. (2011). Enhancing Mathematics Teachers' Knowledge of Students' Thinking from Assessing and Analyzing Misconceptions. *International Journal of Science and Mathematics Education, 1-37.* Springer.

An, S., & Wu, Z. (2014). Using the Evidence-Based MSA Approach to Enhance Teacher Knowledge in Student Mathematics Learning and Assessment. *Journal of Mathematics Education,* Vol 7, No. 2, 108-129

Adamson, F., & Darling-Hammond, L. (2010, April). Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning. Stanford, CA: Stanford Center for Opportunity Policy in Education.

California Department of Education. (2015). *Mathematics framework for California public schools* Sacramento, CA: California Department of Education.

Caffrey, E. D. (2009). *Assessment in Education and Secondary Education: A Primer.*Washington: Congressional Research Service.

Callingham, R. (August 16, 2010). Mathematics assessment in primary classrooms: Making it count. Research Conference 2010. *2009 - 2015 ACER Research Conferences.* Paper 7. Retrived from http://research.acer.edu.au/research_conference/RC2010/16august/7

Hattie, J. A. C. (2009). Visible learning: A synthesis of meta-analyses relating to achievement. Abingdon:
Routledge.

Herman, J.L. & Linn, R.L. (2013). On the road to assessing deeper learning: The status of Smarter Balanced
and PARCC assessment consortia. (CRESST Report 823). Los Angeles, CA: University of California,
National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-
grade arithmetic. American Educational Research Journal, 30, 393-425.
Horn, S., & Sanders, W. (1997). Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education*, 11: 57±67, 1997.
Lewis, A. (2001). Heads in the sand. *Phi Delta Kappan*, 83(1), 3-5.
Marzano, Robert, J. 2006. Classroom Assessment and Grading that Work. ASCD. Alexandria, Virginia.
McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8).
Retrieved April 27, 2016 from http://pareonline.net/getvn.asp?v=7&n=8
Micklo, S.J. (1997). Math portfolios in the primary grades. *Childhood Education*, 73(4), 194-199.
Mok, I.A.C. & Kaur, B. (2006). Learning Tasks. In Clarke, D., Emanuelsson, J., Jablonka, E., and Mok, I.A.C. (eds.) Making Connections: Comparing Mathematics Classrooms Around the World (pp. 147-
200). Rotterdam: Sense Publishers B.V.
National Council of Teachers of Mathematics (NCTM). (1995). Assessment *standards.* Reston, VA: NCTM.
National Council of Teachers of Mathematics (NCTM). (2000). Principles *and Standards for School Mathematics.* Reston: VA: NCTM.
National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). Common Core State Standards for Mathematics. Washington D.C.: National Governors Association
Center for Best Practices, Council of Chief State School Officers.
National Research Council (1993). *Measuring what counts: a conceptual guide for mathematics assessment.*
Washington, DC: National Academy Press.
Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education.* Volume 11, Issue 1, pp 57-67.
Sweet, D. (1993, September). *Performance Assessment*. (Office of Research, Office of Educational Research and Improvement (OERI) of the U.S. Department of Education.) Retrieved from Education Research Consumer Guide: http://www2.ed.gov/pubs/OR/ConsumerGuides/perfasse.html
Smarter Balanced Assessment Consortium. (2012). Claims for the Mathematics Summative Assessment. Retrieved from: http://www.smarterbalanced.org/wp-content/uploads/2015/09/Smarter-Balanced-Mathematics-Claims.pdf
The Nation's Report Card (2013). What level of knowledge and skills have the nation's students achieved? Retrived from http://www.nationsreportcard.gov/reading_math_2013/#/what-knowledge
Thornberg, D. (2001). Pencils down! How decontextualized standardized testing can destroy education. *Multimedia Schools*, 8(3), 6-8.
Tung, R. (2010). *Including Performance Assessments in Accountability Systems: A Review of Scale Up Efforts.* Center for Collaborative Education and the Nellie Mae Education.
Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, DC: The National Academies Press.
Webb, N. (1997). *Research Monograph Number 6: Criteria for alignment of expectations and assessments*
on mathematics and science education. Washington, D.C.: CCSSO.
Wright, A. (2001). The ABCs of assessment. *The Science Teacher*, 68(7), 60-64.
Wu, Z. & An, S. (2007). Using alternative assessments to assess students' learning: Discussion of

traditional and authentic assessments. In N. Idris (Ed.), *Classroom assessment in mathematics education*
(pp. 19-32). Selangor, Malaysia: McGraw-Hill Education.
Wu, Z. (2008). Using the MSA model to assess Chinese sixth graders' mathematics proficiency. *Journal of Mathematics Education. 1*(1), 70-90.
Wu, Z., & An, S. (2015). *Teaching elementary and middle school mathematics u*sing *the MSA approach:*
*Model, strategy, and application (*2nd ed). Irvine, CA: Education for All Publisher.

Game Theory Models for Infectious Diseases

Sheng-He Huang[1], Wensheng Zhou[2], Ambrose Jong[1], Huan Qi[3]
[1]Childrens Hospital Los Angeles, University of Southern California, Los Angeles, CA,
[2]HRL Laboratories, Malibu, CA,
[3]Institute of Systems Engineering, Huazhong University of Science & Technology, Wuhan, China
Corresponding author: shhuang@hsc.usc.edu

Abstract

*The reductionism and Manichaean view of the microbe–human host relationship are the two major limitations in the conventional theories of microbial infection. Three focal point theory-based game models (pure cooperative, dilemma and pure conflict) are proposed for resolving those problems. There exists a dynamic duality relationship (DDR) between symbiosis (Sym) and pathogenesis (Pat) in microbial infection, which is the most fundamental problem in infectomics. Our health is associated with the dynamic interactions of three microbial communities [nonpathogenic microbiota (NP) (Cooperation), conditional pathogens (CP) (Dilemma), and unconditional pathogens (UP) (Conflict)] with the hosts at three different health statuses [nonsusceptibility (NS), conditional susceptibility (CS), and unconditional susceptibility (US)]. Sym and Pat can be quantitated by measuring symbiotic index (**SI**), which is quantitative fitness for the symbiotic partnership, and pathogenic index (**PI**), which is quantitative damage to the symbiotic partnership, respectively. The values of SI and PI can be used for calculation of the symbiotic point (SP).*

1. Introduction

Infectious diseases caused by bacterial, viral, fungal or parasitic pathogens continue to be the leading cause of morbidity and mortality worldwide despite the availability of effective anti-microbial agents and vaccines over the last fifty years [1]. The continual emergence of previously undescribed new pathogens, reemergence of old pathogens, and the rising crisis of antibiotics resistance will certainly heighten the global impact of microbial infections in the 21st century. These problems are mainly due to inadequate knowledge of the dynamic duality relationships (DDR) between symbiosis (Sym) and pathogenesis (Pat) in microbial infections [2]. The term symbiosis, which may have many variations on its definition, in this paper refers to living together through a close and prolonged association between two or more organisms of different species [3, 4]. Duality is defined as different ways of looking at the same thing [5]. There are two major limitations inherent in the conventional theories of microbial infection. On the one hand, in the past century biology and medicine including infectious diseases have been dominated by the reductionistic approaches. Focusing research on individual virulence genes and the important pathogens has been the traditional approach to human infectious diseases. On the other hand, as Joshua Lederberg pointed out [6, 7], medical science is imbued with the Manichaean view of the microbe–human host relationship: “we good; they evil.” Almost all broad-spectrum antimicrobial agents, which are in the best interest of pharmaceutical industries, kill both the good microbes as well as the bad germs. Even though narrow-spectrum anti-infective agents are not “narrow” for pathogens, they also target both the good and bad microorganisms with a limited range of species.

2. Research Questions

Animals and plants are continually infected by an extensive diversity of symbiotic or invading organisms including bacteria, virus, fungus or parasites. Infection of bacteria by phages started long before the emergence of animals and plants [8]. Microbial infection is an evolutionary paradigm which is associated

with co-evolution between hosts and microbes [6, 7, 9]. This co-evolution can be defined as the process of reciprocal and dynamic genetic changes in two or more species [2]. The conventional wisdom in medicine holds that microbial infection is a pathogenic process in which a pathogen enters, establishes itself and multiplies in the host [10]. The emphasis is on the antagonism or conflict, not the mutualism. This represents "zero-sum thinking" – the belief that if one player gains, other player must inevitably lose. Methods and concepts of the zero-sum game theory have proved successful in studying the strategy of pure conflict. The most challenging issue in infectious diseases is how to dissect the dynamic Sym/Pat duality relationships in microbial infections using infectomics and mathematics such as game theory. Game theory, defined in the broadest sense, is the study of the strategies of conflict, cooperation and mixed situations in which both coexist. This article attempts to enlarge the scope and application of game theory in infectious diseases, extending from the zero-sum games to the nonzero-sum games.

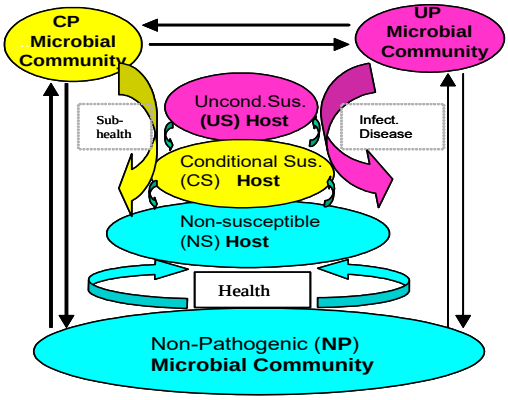


Fig.1 Schematic representation of interactions of three microbial communities [non-pathogenic (NP), conditional pathogenic (CP) and unconditional pathogenic (UP)] with with the hosts at three different health statuses [non-susceptibility (NS), conditional susceptibility (CS), and unconditional susceptibility (US)].

3. Research Evidence and Methods

3.1. Three Community Principles of Microbial Infections

Our health is associated with the dynamic interactions of three microbial communities [2] [nonpathogenic microbiota (NP), conditional pathogens (CP), and unconditional pathogens (UP)] with the hosts at three different health statuses [nonsusceptibility (NS), conditional susceptibility (CS), and unconditional susceptibility (US)] (Fig.1). NP is the major microbial community which forms a healthy symbiotic 'superorganism' with the hosts. The ecology and evolution of NP-NS interaction is essential and fundamental for health. From birth to death, we share a benign coexistence with a vast, complex, and dynamic consortium of microbes. Most of our microbial commensals reside in our gastrointestinal (GI) track packed with up to 100 trillion (10^{14}) microbes [1, 11]. The GI tract harbors a rich microbiota of >600 different bacterial species. Some of these microorganisms have important health functions. These include stimulating the immune system, protecting the host from invading bacteria and viruses, and aiding digestion. The gut microbiota, which is essential for human homeostasis, is established rapidly after birth and remains relatively stable throughout the life [1]. The GI mucosa provides a protective interface between the internal environment and the constant external challenge from food-derived antigens and microbes. CP and UP are minor microbial communities that mainly contribute to the pathogenesis of microbial diseases. The distinction between the commensal and the pathogen in the CP community can be blurred because they may cause diseases under certain sub-health conditions of the hosts, or in immunocompromised hosts. For example, pneumococcus, meningococcus and Haemophilus bacteria regularly exist as part of the normal microbiota of the host respiratory track and are mostly carried asymptomatically despite the fact that they can cause well-defined diseases [12, 13]. Microbes in the CP community dynamically evolve in two opposite directions, which are

toward either the NP (more cooperative or mutualistic) or UP (more pathogenic) microbial community. Microbes with high pathogenicity belong to the UP microbial community. The three microbial communities and three statuses of the hosts are subjected to dynamic reciprocal changes driven by intraspecies, cross-species or cross-kingdom transfer of genetic materials.

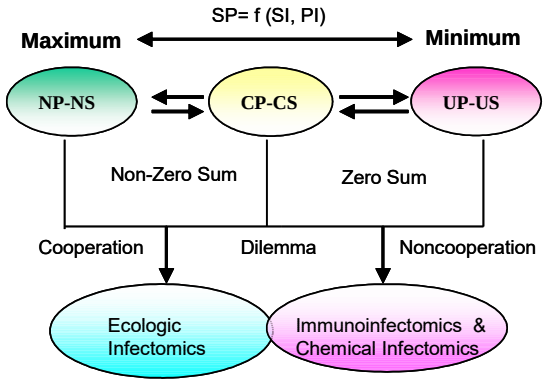


Fig.2. A continuum model of host-microbe interactions coupling with infectomic approaches to dissect the problems in microbial infections.

3.2. Duality Representations of Microbial Infections: Sym and Pat.

Extending along the dynamic continuum from conflict to cooperation, microbial infections always involve symbiosis and pathogenesis, which are two fundamental components of the host-microbe interactions (Fig.2). There exists a Sym-Pat dynamic duality relationship (DDR) in microbial infection, which is the most fundamental issue of infectomics [2]. DDR is reflected in the genotypic and phenotypic infectomes, which are encoded by both host and microbial genomes. The opposition and unity of Sym and Pat are indispensable, and the academic viewpoint that the unity of opposites of Sym and Pat gives impetus to the development of microbial infection is considered as the core idea and radical principle of the duality representations of microbial infections. In certain circumstances and at a certain stage of the development of microbial infection, each of the two aspects of Sym and Pat will transform from antagonism into mutualism or from mutualism into antagonism. Sym and Pat can be quantitated by measuring symbiotic index (**SI**), which is quantitative fitness for the symbiotic partnership, and pathogenic index (**PI**), which is quantitative damage to the symbiotic partnership, respectively. The most crucial experiments are to identify infectomic signatures specific for SI and PI. The set of symbiotic or pathogenic parameters is defined as a function SI(x) or PI(x*). SI(x) and PI(x*) are continuous functions ranging from 0 to 1 to admit different degrees of Sym and Pat, respectively (Fig.3). SI(x)=0 and PI(x*)=0 indicate that x and x* are perceived to be zero-symbiotic and zero-pathogenic, respectively. SI(x)=1 and PI(x*) =1 indicate that x and x* are perceived to be completely symbiotic and completely pathogenic, respectively. Intermediate values of SI(x) and PI(x*) indicate that x and x* are perceived to be partially symbiotic and partially pathogenic, respectively. The hosts have large influences on SI and PI. For example, polydnaviruses have evolved complex life cycles in which they interact as symbionts with one host and pathogens with another. Their genomes reflect the dual roles as mutualists and pathogens [14]. Symbiotic points (SP) are used to determine the DDR between Sym and Pat. The values of SI and PI can be used for calculation of the symbiotic point (SP), which is a function of SI and PI.

SP= f (SI, PI)

The focus of the DDR research is to examine the ability of SP to transform situations of potential conflict (UP-US & CP-CS) into situations of cooperation (NP-NS). SP bears analogy to Schelling's focal point,

which is any feature of such a game that provides a focus of convergence [15]. In the games with multiple Nash equilibria, one equilibrium usually stands out from the others (salient). Such an equilibrium is a focal point which can be easily recognized by all the players [16]. Thomas Schelling's *Strategy of Conflict* (1960) has been recognized as one of the most important works of game theory [17]. There is no doubt that focal points play a central role in Schelling's game theory. Schelling has made a significant contribution to a reorientation of game theory. Understanding focal points is not only a key to improving game theory but also a key to dissecting SPs.

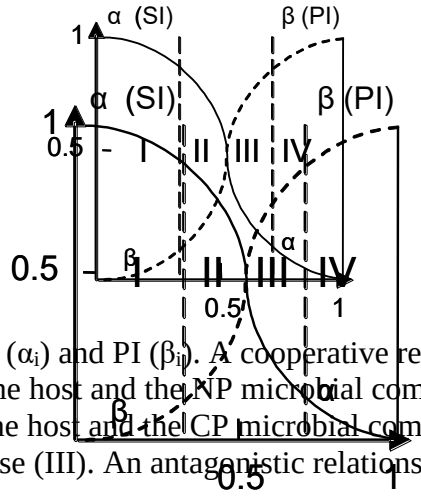


Fig.3. A relationship between SI (α_i) and PI (β_i). A cooperative relationship (NP-NS: health and mutualism) (I) occurs between the host and the NP microbial community. A competitive relationship (CP-CS) (II and III) exists between the host and the CP microbial community. There are two types of competitions, better (II) and worse (III). An antagonistic relationship (UP-US) (IV) between the host and the UP microbial community.

3.3. Game Theoretical Models (GTMs) of Microbial Infections.

In this paper, three types of GTMs are proposed for studies on NP-NS interactions (cooperative game), UP-US interactions (noncooperative games), and CP-CS interactions (dilemma or bargaining game). First, the NS-NS interactions are dissected with pure cooperative games in which each player chooses the strategy corresponding with the focal point in the expectation that the others will do the same. The significance of focal points can be shown most clearly in the pure cooperative games (Fig.4(a)). As there is no conflict of interests in these games, all the players merely want to cooperate and they do not choose the alternative ways. Analysis of the cooperative game issues is to focus on coalition formation and distribution of the gains through cooperation. The SP in the NP-NS games tends to be maximal (Fig.2). Secondly, noncooperative GTMs are used for analysis of the UP-US interactions. In contrast to cooperative games which focus on collective rationality and common interest, noncooperative games emphasize individual rationality and individual optimal strategy. The SP in the UP-US games tends to be minimal (Fig.2). In games of pure conflict, defection is the equilibrium strategy and the total benefit to all players in the game, for every combination of strategies, always adds to zero (zero-sum). This situation is depicted in Fig.4(b) as a two-player game. In the antagonistic UP-US interaction model, the surviving strategies of the UP community conflict with that of the US host. The UP evolves to exploit the host as much as possible, and the host adapts to exclude or limit the damage caused by the UP. Thirdly, we consider the strategic use of focal point theory in mixed situations to analyze the CP-CS interactions in which there is both conflict and mutual dependence. The most well-known example is the Prisoner's Dilemma game (a two player game) in which each player chooses between a cooperating and defecting strategy. As shown in Fig.4(c), each player receives a higher playoff by defecting than by cooperating, no matter what the other player chooses. However, they receive a higher playoff if both cooperate than both defect. The two player game can be extended to the N-player Prisoner's Dilemma game with arbitrary numbers of players. The CP-CS interactions can co-evolve toward two different directions, increasing or decreasing the SP.

4. Outcome

In this paper, focal point theory-based game models are proposed for analysis of the dynamic duality relationships (DDR) of Sym-Pat in microbial infections. DDR is the most fundamental problem in infectomics, which is the integration of omics and mathematical/computational approaches. There are three types of infectomic approaches that can be used for the control of microbial infections: ecological infectoimcs, immunoinfectomics and chemoinfectomics [2]. Ecological infectomics will explore symbiotic solutions to microbial infections. Developing novel immunological intervention strategies for the prevention and treatment of microbial infections using infectomic signatures and immunomic approaches falls within the field of immunoinfectomics. Chemoinfectomics represents the most powerful approach to the development of a new generation of drugs for antimicrobial chemotherapy.

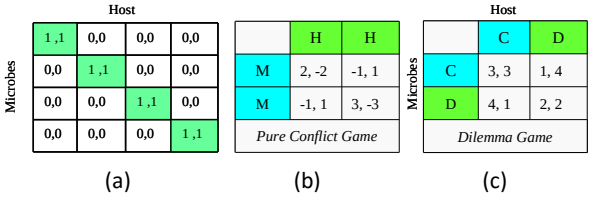


Fig.4. (a) Pure cooperative game. (b) Pure conflict game. H: host; M: microbes. (c) Dilemma game. The number in the left of each pair indicates the payoff for Microbes; the right, Host. Higher numbers represent greater payoff for the individual. Two strategies [Cooperation (C) and Defection (D)] are used.

4.1. Symbiosis Point Converting (SPC): Ecological infectomics-based approaches for rational control of microbial infections. As microbial infection is an ecological and evolutionary paradigm which is associated with co-evolution between hosts and microbes (such as human host and microorganisms, phages and bacteria) in dynamic ecosystems, two ecological infectomics-based SPC approaches (increasing and decreasing SP) can be used for rational control of infectious diseases [2]. The focus in SP increasing approaches is how to transform situations of potential conflict (pathogenesis) into cooperation (symbiosis) by dissecting the dynamic duality relationships (DDR) between Sym and Pat in microbial infections and developing symbiotic agents (symbiotics) that favor a healthy symbiosis [2]. Symbiotics are defined as products that are beneficial to symbiotic ecology of the super-organisms consisting of microbes and their human hosts. These include microbial (e.g., probiotic bacteria) and nonmicrobial agents (e.g., prebiotics) [2]. The introduction of beneficial symbiotics with higher SP in our body should be a very attractive rationale for modulating the microbiota, improving the symbiotic homeostasis of the superorganism and providing a microbial stimulus to the host immune system against pathogens. Decreasing SP is another rational strategy for control of microbial infections. As phages, which specifically kill bacteria, play an important role in the ecology, evolution and virulence of a number of pathogens, there is a rational use of phages for treatment and prevention of bacterial infections. The use of phages to treat bacterial infections has a long history dating back to mid 1910's [2]. Due to the availability of effective broad spectrum antibiotics in the early 1940's, phage therapy was discarded in Western medicine at that time. The rising crisis of antibiotic resistance has recently increased great interest in phages and their use as natural antimicrobial agents to fight microbial infections [2]. Compared with commonly used antibiotics, a great advantage of phages is their narrow host range. Recent studies showed that co-infection with GB virus C (GBV-C) is associated with a decreased mortality in HIV-infected patients [18]. Therefore, reducing SP between microbial agents (such as phages and GBV-C) and targeted pathogens is another excellent ecological approach for the development of novel antimicrobial agents.

4.2. Specific Pathogen-Targeting (SPT): Immunoinfectomics- and chemoinfectomics-based approaches for prevention and treatment of infectious diseases. In contrast to the ecological infectomics-based SPC approaches that focus on the symbiotic relationships (such as NP-NS and CP-CS

interactions) between the hosts and microbial communities, immunoinfectomics- and chemoinfectomics-based SPT approaches emphasize the use of antagonistic relationships (such as UP-US interactions) between the hosts and microorganisms. It is important to point out that the SPT approaches are intrinsically different from the conventional pathogen-targeting antimicrobial agents, which kill both pathogens and nonpathogens [2]. The availability of the genomic information from both microbes and their hosts has resulted in exciting new progress in the field of immunoinfectomics. Nanobody-based technologies and immune epitope mapping have emerged as the very powerful tools for the discovery and development of novel antimicrobial agents. Concurrent advances in both high-throughput chemistry and infectomics have given rise to the field of chemoinfectomics for elucidating and validating drug targets, and generating novel therapeutics. Chemoinfectomics refers to the use of small synthetic molecules that are highly specific for defined infectomic targets, for biological function analysis and to discover new drug leads. The progress towards understanding the DDR of Sym-Pat in microbial infections using focal point theory-based game models will greatly facilitate the use of ecological infectomics, immunoinfectomics and chemoinfectomics for the rational control of infectious diseases.

5. References

[1]. S.H. Huang, T. Triche and A.Y. Jong, "Infectomics: genomics and proteomics of microbial infections," Functional & Integrative Genomics. Vol.1, pp.331-44, 2002.
[2]. S.H. Huang, X. Wang and A. Jong, "The evolving role of infectomics in drug discovery," Expert Opinion in Drug Discovery, Vol. 2, pp.1-15, 2007.
[3]. M.J. Roossinck, "Symbiosis versus competition in plant virus evolution," Nature Review Microbiology, Vol. 3, pp.917-24, 2005.
[4]. G.G. Dimijian, "Evolving together: the biology of symbiosis," Baylor University Medical Center Proceedings, Vol. 13, pp.217-26, 2000.
[5]. S. Lee, "Three roads to Quantum Gravity," Basic Books, 2001, page 113.
[6]. J. Lederberg, "Infectious history," Science, Vol. 288, pp.287-93, 2000.
[7]. J. Lederberg. "The future of infectious diseases," Journal of Urban Health. Vol.75, pp.463-70, 1998.
[8]. H. Brussow, C. Canchaya, W.D. Hardt, "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion," Microbiology and Molecular Biology Reviews, Vol. 68, pp.560-602, 2004.
[9]. M.E. Woolhouse, J.P. Webster, E. Domingo, B. Charlesworth, B.R. Levin, "Biological and biomedical implications of the co-evolution of pathogens and their hosts," Nature Genetics, Vol.32, pp.569-77, 2002.
[10]. A. Casadevall and L.A. Pirofski, "Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease," Infection and Immunity, Vol. 68, pp.6511-8, 2000.
[11]. R.E. Ley, D.A. Peterson, J.I. Gordon, "Ecological and evolutionary forces shaping microbial diversity in the human intestine," Cell. Vol.124, pp.837-48, 2006.
[12]. S. Falkow, "Is persistent bacterial infection good for your health?" Cell, Vol.124, pp.699-702, 2006.
[13]. D. Kuklinska, M. Kilian, "Relative proportions of Haemophilus species in the throat of healthy children and adults," European Journal of Clinical Microbiology, Vol.3, pp.249-52, 1984.
[14]. B.A. Webb, M.R. Strand, S.E. Dickey, et. al., "Polydnavirus genomes reflect their dual roles as mutualists and pathogens," Virology, Vol. 347, pp.160-174, 2006
[15]. K. Benmore, L. Samuelson, "The evolution of focal points," Games and Economic Behavior, Vol.55, pp.21-42, 2006
[16]. J. Metha, C. Starmer, R. Sugden, "Focal points in pure coordination games: an experimental investigation," Theory and Decision, Vol. 36, pp.163-185, 1994.
[17]. R. Sugden, I.E. Zamarron, "Finding the key: the riddle of focal points," Journal of Economic Psychology, Vol. 27, pp.609-621, 2006

[18]. D.E Yirrell, E. Wright, L.A. Shafer, et. al., “ssociation between active GB virus-C (hepatitis G) infection and HIV-1 disease in Uganda,” International Journal of STD & AIDS, Vol. 18, pp.244-9, 2007.

*Published in Frontiers in the Convergence of Bioscience and
Information Technologies, 2007. FBIT 2007
Date of Conference: 11-13 Oct. 2007

Clinical Course of Chronic Hepatitis B (CHB) Presented with Normal ALT in Asian American Patients

Kelvin Nguyen[1*], Calvin Pan[2*], Victor Xia[3], Jennifer Hu[2], Ke-Qin Hu[1]

[1]Division of GI/Hepatology, University of California, Irvine, School of Medicine, Orange, CA, United States; [2]Division of Gastroenterology and Hepatology, NYU Langone Medical Center, NYU School of Medicine, NY; [3]Dept of Anesthesiology, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

ABSTRACT

BACKGROUND: *The clinical course for chronic hepatitis B (CHB) patients with normal ALT and with or without minimal histologic activity remains unclear.*
GOAL: *We assessed frequency, amplitude, disease activities, and associated factors of ALT and/or AST flares in this subpopulation.*
STUDY: *47 consecutive treatment naïve Asian CHB patients were enrolled from two liver clinics between December 2003 and January 2013, who had normal baseline ALT by routine clinical biochemical testing done 6 weeks before or after the liver biopsy (LB). Flare is defined as elevation of ALT/AST above ULN of ALT/AST.*
RESULTS: *The mean post-LB follow-up was 37.6 (C.I. = 12, 88) months. The mean age at the LB was 43.3 (C.I. = 19, 65); 22/47 (46.8%) were males; 15/45 (33.3%), HBeAg+; 68.1% had stage 0-1 fibrosis; 63.8% had grade 0-1 inflammation. During follow-up, 13/47 (27.7%) cases developed ALT flare at least once in a mean of 13.5 (C.I. = 2, 43) months after LB; ALT flare was not associated with baseline ALT level, fibrosis stage, inflammation grade, hepatitis B virus (HBV) DNA load, HBeAg status, HBV genotype, HBV pre-core, and basal core promoter mutations. 11/13 (84/6%) of ALT flares resolved during follow-up. 13/13 (100%) of ALT flares met AASLD treatment criteria but only 6/13 (46.2%) were on HBV treatment.*
CONCLUSIONS: *Serum ALT and/or AST flares occur frequently in CHB carriers who initially presented with normal ALT during pre-treatment period. Thus, regular follow-up is warranted despite status of ALT/AST. No clinical factors were found to be associated with ALT flares.*
Key Words: *Chronic hepatitis B; Normal ALT; Liver biopsy; ALT flare; AST flare*

INTRODUCTION

Chronic hepatitis B virus (HBV) infection affects more than 400 million people worldwide. Vast majority of chronic hepatitis B (CHB) patients are located in Asia. (1) CHB causes ongoing liver injury that can be associated with cirrhosis and hepatocellular carcinoma (HCC). Compared to people without HBV infection, HBV carriers have a 100 fold higher relative risk of developing HCC. (2) It is estimated that 15-40% of CHB patients will progress to cirrhosis that can be complicated by hepatic decompensation and end stage liver disease, while 25% of these patients will develop HCC. (2-5) The rate of developing cirrhosis in patients with chronic HBV infection at tertiary centers ranges from 2-7%

annually. [6-10] In order to evaluate the severity of liver injury, liver biopsy remains the gold standard modality to accurately evaluate fibrosis/cirrhosis in patients with CHB. [11]

Serum ALT level has been used as one of the parameters in determining clinical phases of HBV infection and serum ALT elevation, also called ALT flare, is an indication of hepatic necroinflammation. [12-13] Acute exacerbation, manifested with ALT flare, occurs in patients with CHB, especially in those with elevated ALT levels at presentation. [14] Most of these exacerbations were associated with reactivation of HBV replication and ended with spontaneous resolution and ALT normalization, though the exacerbation could recur.

Although CHB patients with persistently normal ALT are thought to have mild hepatic necroinflammatory activity on liver biopsy and have good prognosis, [15-16] several studies have demonstrated that CHB patients with significant inflammation and fibrosis on liver biopsy can present with normal or minimally elevated ALT, questioning the accuracy of current standard ALT criteria in reflecting HBV-related liver injury. [17-25]

Serum ALT is an important parameter used in all major guidelines for HBV treatment from different organizations. [26-29] In general, the decision to initiate HBV treatment is based on CHB natural history and possible risks for disease progression. Both Asian Pacific Association for the Study of the Liver (APASL) and American Association for the Study of Liver Diseases (AASLD) recommend treatment once the patients' ALT is above 2 x Upper Limit of Normal (ULN), while European Association for the Study of the Liver (EASL) and Keffee and Tong's algorithms recommend to consider treatment once ALT is more than 1 x ULN [12, 26-29] as studies have shown that when ALT gets close to ULN, the patients may potentially have increased liver related mortality. [30-31] On the other hand, it is well known that single elevations of ALT level might not indicate significant severe histologic disease. [22-23, 32] However, it is still unclear about the clinical course of CHB patients with normal ALT with or without minimal histologic activity. A good understanding about disease progression will facilitate the discussion when it is appropriate to start treatment for CHB patients with normal ALT and AST at baseline.

In the present study, we evaluate CHB patients with normal ALT and AST at baseline for frequency of HBV infection flare, specifically ALT and AST flare from the time of liver biopsy. We also assess risk factors of ALT flare in these Asian American patients with CHB.

PATIENTS AND METHODS

Patients and Enrollment Criteria. This study was conducted via a consecutive and retrospective chart review of patients with CHB who were regularly followed in 2 outpatient liver clinics at University of California, Irvine medical center and Elmhurst Hospital of Mount Sinai School of Medicine between December 2003 and January 2013. The Institutional Review Board at both centers approved this study and the informed consent was exempt. Patients were enrolled if they met the following inclusion criteria: Asian American patients who carried diagnosis of CHB with detectable serum HBV DNA; underwent a liver biopsy with report of normal or mild liver histology i.e. METAVIR Staging ≤ 2 and Fibrosis ≤ 2; regular follow-up every 6 months at minimum with liver function tests and HBV DNA for minimal 12 months; and serum ALT level $\leq$ ULN. The ULN of ALT is 60 IU/L at University of California, Irvine liver clinic while the ULN is 40 IU/L at Mount Sinai clinic. Exclusion criteria include coinfection with HIV, Hepatitis C or D, or HCC. A total of 47 patients who met above inclusion criteria were included in the present study.

Clinical Data Collection. Chart review was performed to collect clinical information that includes date of birth, age, gender, ethnicity, date of clinic entrance, date of biopsy, date of treatment

initiation if applied. Follow up time started from the time the patient had entered the clinic till January 2013 or the time that treatment was started. Once a patient received treatment for hepatitis B, that patient's clinical data will no longer be followed by our study.

Laboratory Data Collection. The following laboratorial data were collected at the time of liver biopsy as the baseline results, then every 3-6 months with a two-week variation. Biologic and hematologic laboratory data included ALT, AST, total bilirubin, albumin, protime/INR, and platelet count. HBV serological and virological test results included HBsAg, HBeAg, HBeAb, HBV DNA level, HBV genotype, pre-core (PC) mutations (G1896A and C1858T), and basal core promoter (BCP) mutations (A1762T and G1764A). ALT ≤ 60 IU/L is considered within normal limit (WNL) at University of California, Irvine liver clinic whereas ALT ≤ 40 IU/L is the reference for WNL at Mount Sinai clinic. AST ≤ 40 IU/L is WNL at both clinics.

Pathology Data Collection. All percutaneous liver biopsies were performed at our centers. Liver histology was assessed by pathologists specialized in liver diseases without any knowledge of the clinical data. A liver sample was considered adequate if it was longer than 15 mm and contained six portal tracts or more. The METAVIR scoring system was used to evaluate histologic grades and stages of the liver biopsy. [33-35] The severity of necroinflammation was graded from A0 (no activity) to A3 (severe activity) and the stage of fibrosis was graded from F0 (no fibrosis) to F4 (cirrhosis). [35]

Statistical Analysis. The descriptive data were shown as whole numbers and percentages. Chi-square tests were performed in order to evaluate univariate analysis of predictive risk factors for ALT flare. Student t-tests were also performed. A p-value of less than 0.05 is considered statistically significant. The primary endpoints were frequency of ALT/AST flares and associated factors with ALT flare in this subpopulation of patients with mild histologic disease activity and normal liver enzymes at the time of biopsy.

RESULTS

Baseline Characteristics of the Patients. As summarized in Table 1, of 47 patients in the study, all were of Asian American background and had biopsy with ALT ≤ ULN, specific for each institution's laboratory reference. Mean follow-up was 37.6 ± 19.0 months with the range of 12-88 months. Mean age was 43.3 ± 13.5 years with 22/47 (46.8%) being male. Mean ALT at biopsy was 26.5 ± 10.3 IU/L, while mean AST at biopsy was 24.6 ± 6.7 IU/L. Mean HBV DNA level at biopsy was $2.0 \times 10^8 \pm 1.0 \times 10^9$ IU/mL. All patients with liver biopsy reported histologic stage and grade ≤ 2, 68.1% and 63.8% had fibrosis stage ≤ 1 and inflammatory grade ≤ 1, respectively. Of 45 patients, 33.3% were HBeAg positive, 71.1% developed HBeAb; 2 cases were positive for both HBeAg and HBeAb. Ninety-eight percent of the participants had genotype B or C, 66.7% had genotype B infection. HBV PC mutations accounted for 62.8% of all cases, HBV BCP mutations accounted for 32.6% of cases, while 20.9% of all cases had both PC and BCP mutations, i.e. at least one PC mutation (G1896A or C1858T) and one BCP mutation (A1762T and G1764A).

Frequency and Clinical Features of CHB Flares. Of 47 participants, 13/47 (27.7%) were found to have ALT flares defined as ALT > ULN (Figure 1), 10/13 (76.9%) had a single flare; 2/13 (15.4%), 2 flares; and 1/13 (7.7%), 5 flares during follow-up. In these ALT flare events, 10/13 (76.9%) were associated with AST flares. The first episodes of ALT flare occurred between 2 and 43 months after liver biopsy with a mean interval of 13.5 ± 13.3 months. The mean ALT of these flares was 114.5 ± 107.3 IU/L. Only 1/13 (7.7%) case with first ALT flare was associated with further HBV DNA increase by more than 1 log of 10. In 11/13 (84.6%) patients, the first episodes of ALT flare resolved eventually, in 19.7 ± 20.0 weeks. However, only 5/13 of those cases experienced spontaneous resolution of ALT flare i.e. resolving within 12 weeks. The episodes of ALT flare achieved the highest ALT at a mean

value of 148 ± 116.1 IU/L, after a mean interval of 21.2 ± 15.7 months since biopsy, and resolved after a mean interval of 17.8 ± 16.4 weeks, ranging from 2 to 52 months. (Table 2)

Among 12 (25.5%) cases with first episode of AST flare, 2/12 (16.7 %) were AST flare only and not accompanied with ALT flare. The first episode of AST flare reached a mean AST of 79.0 ± 86.2 IU/L, ranging from 39 to 344 IU/L. These episodes occurred at a mean interval of 17.2 ± 16.2 months and resolved within 19.7 ± 20.0 weeks. The AST flares with the highest AST level were obtained at a mean AST of 95.4 ± 85.7 IU/L, varying from 41 to 344 IU/L. They occurred after an interval of 24.5 ± 17.1 months since liver biopsy, and resolved in 15.7 ± 15.9 weeks, ranging between 2-52 weeks. (Table 2)

Outcomes and Factors Associated with CHB Flares. During the followup, none of these 13 cases with ALT flare or 12 cases with AST flare experienced hyperbilirubinemia, hepatic decompensation, or required hospitalization. As shown in Figure 1, in 13 patients with ALT flare, 6 (6/13, 46.2%) met criteria of AASLD recommendations [27] and were started on HBV treatment, and 7 (7/13, 53.8%) met the criteria of AASLD recommendations, but chose not to undergo HBV treatment. In these 7 non-treated patients, 3 cases (42.9%) had spontaneous resolution of ALT flare, defined as ALT returning to normal range within 12 weeks. However, 2/7 (28.6%) non-treated cases with ALT flare experienced recurrent ALT flare after a mean period of 17.5 months since resolution of the first flare.

As shown in Table 4, the univariate analysis has shown no significant association between ALT flare and all assessed clinical factors, including age ≥ 50, gender, ALT at biopsy ≥ ½ ULN, fibrosis stage > 1, inflammation grade > 1, HBV DNA ≥ 100,000 IU/mL or ≥ 20,000 IU/mL, HBeAg positivity, HBV PC mutation, HBV BCP mutation, any combination of PC and BCP mutations, and genotype B or C.

DISCUSSION

Although biochemical flare is well known as part of natural history of CHB, the outcomes of ALT flare remains to be determined in patients with normal baseline ALT and AST and mild histologic disease activity on liver biopsy. In the present study, a good number of patients developed flares, especially ALT flares (27.7%). Amongst patients with ALT flares, most (76.9%) have both ALT and AST flares. Most episodes of ALT flare are single flares, except one case with five flares, which was the highest among all flares. Although most of these flares resolved eventually, 2/13 (15.4%) cases of ALT flares did not show flare resolution in a mean follow-up of 24 (0-48) months since the flares occurred and probably due to lack of closer and longer follow-up.

HBV-related biochemical flare can be classified as either host induced, virus induced, or indeterminate. A host induced flare could be a signal of HBeAg seroconversion, whereas a virus induced flare means either the virus is replicating rapidly and causing liver damage or reactivation of HBV or reversion from anti-HBe to HBeAg. [36-42] An ALT flare with elevated HBV DNA level indicates significant infectivity with potential liver damage, prompting immediate hepatitis B treatment. Biochemical flares in our study are mostly not associated with elevations in HBV DNA levels, which deemed most of the flares in our patient population as host-induced. It should be noted, as a retrospective study, we could not rule out other possible non-HBV causes for ALT flare, such as alcohol use, other virus infection, medications, or even herbal supplements.

Is HBV DNA a better marker than ALT for liver injury? As mentioned by Iloeje *et al*. in the REVEAL HBV study, risk for cirrhosis increases significantly with increasing HBV DNA levels and is independent of HBeAg status and serum ALT. [43] Most of the flares in our study were not associated with more than one log increase in HBV DNA level. Only one flare was

observed to have an increase by three logs of 10 in HBV DNA viral load. This observation shows there is no clear association between HBV DNA elevation and biochemical flares in CHB patients in our cases. As long-term changes in serum levels of HBV DNA and ALT are independent predictors for HCC, regular monitoring of HBV DNA levels and ALT is important in clinical management of chronic carriers of HBV. (44)

Currently, no simple and readily available markers were able to predict biochemical flares accurately in CHB patients. We assessed potential risk factors for ALT flares, including gender, age, ALT and HBV DNA levels at biopsy, histology of liver biopsy, HBeAg status, HBV PC and/or BCP mutations, and genotypes. However, we found no significant correlation between these factors and ALT flares. A larger patient population study will be needed to confirm our findings. Although no flare associated factor was identified, this study has shown that CHB patients with significant viremia, even with normal ALT and minimal or normal liver histology, will have significant risk of disease activity during the follow-up period. Thus, close monitoring is needed for this population of patients.

In this cohort of patients, only 46.2% of patients with ALT flare were treated with antiviral therapy. The continuous follow-up on the natural history of ALT/AST flare had to be interrupted to comply with the good clinical practice. However, majority of patients reviewed in this study with ALT/AST flare did not enter into the treatment. Our observation pointed to the direction of a treatment barrier from both physicians and patients due to the lack of recognition on the impact of ALT flare. Therefore, further physician and patient education is needed on HBV monitoring for treatment candidate selection.

One of the main reasons that patients met criteria for treatment, but not receiving treatment is due to patient's refusal. It is difficult to have patients convinced on starting HBV treatment when most of patients with flares are asymptomatic. (12-13, 45) In addition, most of the patients with flares had one flare only and their ALT elevation resolved quickly, evidenced by normal follow-up liver function tests. This creates a false alarm that the disease rarely exacerbates and even if it flares up, it will become quiescent quickly so that close monitor is an acceptable option.

Of all 12 cases of AST flare, 6 (50%) cases received HBV treatment. 2/12 AST flares were AST flare without ALT flare and only one of these two cases underwent HBV treatment due to significant high HBV DNA level in the billion IU/mL range. AST has not been used as one of the criteria considered for HBV treatment. However, all of the cases of AST flare have met AASLD treatment criteria, based on ALT and HBV DNA viral load. Therefore, the utility of AST as one of the treatment criteria is still questionable.

According to findings of our study, patients with CHB, regardless HBeAg status i.e. HBeAg-negative or HBeAg-positive, are at risk for ALT flares. With the emergence of more updated, less stringent guidelines for HBV treatment, we are anticipating more patients will receive early treatments. In the mean time, frequent and close follow-ups are very important.

Regarding surveillance for HCC in Asian Americans with hepatitis B, the common guideline consists of alpha fetoprotein (AFP) and abdominal ultrasound every 6 months. (12) Surveillance candidates are high-risk patients who are at advanced stages of liver disease such as cirrhosis, HCC in blood relatives, or low to moderate risk patients such as inactive carriers, immune tolerant patients, HBsAg positive males < 40 years old, HBsAg positive females < 50 years old or patients with HBsAg loss. (12) Patients in our clinic are routinely seen every 3-6 months and we would recommend these frequent follow-ups.

HBeAg seroconversion indicates immune clearance, which reduces risk of hepatic decompensation and improves survival. (46-47) However, there is a concern that up to 30% of CHB

patients continue to have elevated ALTs and high HBV DNA levels after HBeAg seroconversion, [37, 48-49] making them more prone to develop complications of liver cirrhosis. [6, 39, 50] 66.7% of our study participants are HBeAg negative. Of 12 patients with ALT flares and known HBeAg status, 8 (66.7%) patients were tested as HBeAg negative and 6/7 (85.7%) of them had high level of HBV DNA (i.e. more than 100,000 IU/mL). This finding is critical as the association between HBV DNA and HCC is greater in the subset of patients with HBeAg negative with normal ALT levels and no cirrhosis at entry. [51] Therefore, these patients should be highly considered for HBV treatment as well as continuing close follow-ups.

There are a few limitations in this study. We were not able to identify any clinical factors associated with ALT flare in our study, possibly due to a low power effect. A larger patient population might be able to show a statistically significant correlation between some of these factors and ALT flare. Despite a negative result, our study has established a foundation for future studies to look into the natural disease progression of CHB as well as risk factors for serologic flares. According to Prati et al, ULN for ALT should be 30 IU/L for men and 19 IU/L for women as patients with higher levels of ALT than the above suggested values and less than 40 IU/L, which was traditionally the common ULN, were found to have histologic disease activities. [30]

Our study includes patients from 2 different clinics where the reference ranges for ALT and AST are not the same. Therefore, a subset analysis for ALTs lower than the common ULN by Prati standard was not performed.

In conclusion, our study has shown a significant proportion of CHB patients with ALT/AST ≤ ULN developed flares during follow-up after biopsy. These biochemical flares were not associated with worsening HBV DNA level and were not associated with seroconversion.
The fact that many of these patients with flares did not receive treatment brought attention to closer follow-ups on these patients. Threshold for treatment in CHB patients should be lower than it is now and early treatment for these patients is highly recommended. So far, no risk factors have been identified as being associated with ALT flare.

ACKNOWLEDGEMENTS
An early analysis of the study was presented at the 2012 AASLD meeting in Boston.

DISCLOSURE

Dr. Ke-Qin Hu has received research or educational grants from Gilead, Bristol Myers Squibb, Genentech, and Merck. He also serves as speaker bureau for Gilead, Bristol Myers Squibb, Genentech, and Merck.
Dr. Calvin Q. Pan has received research grants from Gilead, Bristol Myers Squibb, Novartis, and Roche. He also serves as a consultant, advisor and speaker bureau for Gilead and Bristol Myers Squibb.

Table 1. Baseline Clinical Presentation

Variables	Cases	Percentage or Mean ± SD
Age (years old)	47/47	43.3 ± 13.5
Gender (male)	22/47	46.8%
Post-Biopsy follow up (months)	47/47	37.6 ± 19.0
ALT at biopsy (IU/L)	47/47	26.5 ± 10.3
AST at biopsy (IU/L)	47/47	24.6 ± 6.7
HBV DNA (IU/mL)	45/47	476 – 1.1 x10^9
HBV DNA log at biopsy	45/47	5.0 ± 2.0
Histologic stage ≤ 1	32/47	68.1%
Histologic grade ≤ 1	30/47	63.8%
No fatty liver on histology	39/44	88.6%
Genotype B	28/42	66.7%
Genotype C	13/42	31.0%
HBeAg+ at biopsy	15/45	33.3%
HBeAb+ at biopsy	32/45	71.1%
Precore mutation (G1896A or C1858T)	27/43	62.8%
HBeAg+	3/26	11.5%
HBeAg-	23/26	88.5%
BCP mutation (A1762T or G1764A)	14/43	32.6%
HBeAg+	5/14	35.7%
HBeAg-	9/14	64.3%
Pre-core AND BCP mutations	9/43	20.9%
HBeAg+	1/9	11.1%
HBeAg-	8/9	88.9%

Table 2. Clinical Presentation and Clinical Course after Liver Biopsy

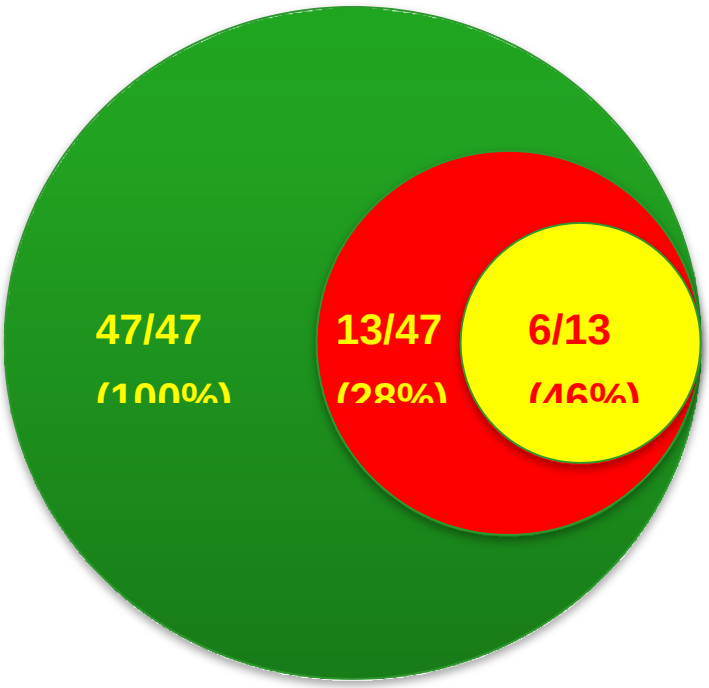
Variables	Cases	Percentage or Mean ± SD
Patterns of ALT flares		
All with ALT flares	13/47	27.7%
Cases of single ALT flares	9/13	69.2%
Cases of 2 ALT flares	3/13	23.1%
Cases of 5 ALT flares	1/13	7.7%
All cases of AST flares	12/47	25.5%
Cases with only AST flares	2/12	4.3%
Cases of ALT flares associated AST flares	10/13	76.9%
Presentation of ALT flares		
ALT flares and HBeAg status		
HBeAg-positive	4/12	33.3%
HBeAg negative	8/12	66.7%
First ALT flare (IU/L)	13/47	114.5 ± 107.3
Biopsy to first ALT flare (months)	13/47	13.5 ± 13.3
Recovery from first ALT flare (weeks)	11/13	21.3 ± 19.1
HBV DNA at first ALT flare (IU/mL)	11/47	$7.5x10^7 + 2.3x10^8$
Flare with highest ALT (IU/L)	13/47	148 ± 116.1
Biopsy to flare with highest ALT (months)	13/47	21.2 ± 15.7
Flare with highest ALT resolves in (weeks)	10/47	17.8 ± 16.4
First AST flare (IU/L)	12/47	79.0 ± 86.2
HBV DNA at first AST flare (IU/mL)	11/47	$1.5x10^7 + 3.1x10^7$
Biopsy to first AST flare (months)	12/47	17.2 ± 16.2
First AST flare resolves in (weeks)	11/47	19.7 ± 20.0
Flare with highest AST (IU/L)	12/47	95.4 ± 85.7
Biopsy to flare with highest AST (months)	12/47	24.5 ± 17.1
Flare with highest AST resolves in (weeks)	10/47	15.7 ± 15.9

Table 3. Summary of the Outcomes from CHB Biochemical Flares

Outcomes	Cases	Percentage
HBV Tx started in general	15/47	31.9%
HBV Tx started with ALT or ALT/AST flares	6/15	40.0%
HBV Tx started with AST flares only	1/15	6.7%
% of ALT flares that met Rx criteria	13/13	80.0%
% of ALT flares, meeting Rx criteria and started on Rx	6/13	46.2%
% of ALT flares without treatment	7/13	53.8%

Table 4. Univariate Analysis of Possible Risk Factors for CHB Biochemical Flares

Variables	Cases/Total ALT flares (%)	P-value
Age ≥ 50	17/47 (36.2%)	0.3
Male gender	8/13 (61.5%)	0.3
ALT at biopsy ≥ ½ ULN	10/13 (76.9%)	0.3
Fibrosis stage > 1	9/13 (69.2%)	1.0
Inflammation grade > 1	8/13 (61.5%)	1.0
HBV DNA ≥ 20,000 IU/mL	12/12 (100.0%)	0.2
HBV DNA ≥ 100,000 IU/mL	10/12 (83.3%)	0.2
HBeAg positivity	4/12 (33.3%)	1.0
Any combined precore/BCP mutations	2/12 (16.6%)	1.0
Genotype B	10/12 (83.3%)	0.3
Genotype C	2/12 (16.7%)	0.3
Genotype B or C	12/12 (100.0%)	1.0
Precore mutation G1896A only	4/12 (57.1%)	1.0
Any precore mutation (G1896A or C1858T)	6/12 (50.0%)	0.5
BCP mutations (both A1762T AND G1764A)	3/7 (42.9%)	0.1
Any BCP mutation (A1762T OR G1764A)	3/12 (25.0%)	0.7
G1896A with both A1762T and G1764A mutations	2/7 (28.6%)	0.2
Any combination of precore and BCP mutations	2/12 (16.7%)	1.0



Total number of cases

Cases with ALT flare

Cases with ALT flare and HBV treatment started

Figure 1. Distribution of ALT flare and HBV treatment

References

1. Custer B, Sullivan SD, Hazlet TK, *et al*. Global epidemiology of hepatitis B virus. *J Clin Gastroenterol* 2004; **38**: S158-68.
2. Beasley WM. Hepatitis B virus-the major etiology of hepatocellular carcinoma. *Cancer* 1988; **61**: 1942-56.
3. Lee WM. Hepatitis B virus infection. *N Engl J Med* 1997; **337**: 1733-45.
4. Lok AS and McMahon BJ. Chronic hepatitis B. *Hepatology* 2007; **45**: 507-39.
5. Sanchez-Tapias JM, Costa J, Mas A, *et al*. Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in Western patients. *Gastroenterology* 2002; **123**: 1848-56.
6. Hsu YS, Chien RN, Yeh CT, *et al*. Long-term outcome after spontaneous HBeAg seroconversion in patients with chronic hepatitis B. *Hepatology* 2002; **35**: 1522-27.
7. Fattovich G, Brollo L, Guistina G, *et al*. Natural history and prognostic factors for chronic hepatitis type B. *Gut* 1991; **32**: 294-98.
8. Huo T, Wu JC, Hwang SJ, *et al*. Factors predictive of liver cirrhosis in patients with chronic hepatitis B: a multivariate analysis in a longitudinal study. *Eur J Gastroenterol Hepatol* 2000; **12**: 687-93.
9. Lin SM, Sheen IS, Chien RN, *et al*. Long-term beneficial effect of interferon

therapy in patients with chronic hepatitis B virus infection. *Hepatology* 1999; **29**: 971-75.

10. Liaw YF, Tai DI, Chu CM, *et al*. The development of cirrhosis in patients with chronic type B hepatitis: a prospective study. *Hepatology* 1988; **8**: 493-96.

11. Wai CT, Cheng CL, Wee A, *et al*. Noninvasive models for predicting histology in patients with chronic hepatitis B. *Liver Int.* 2006; **26**: 666-72.

12. Tong MJ, Pan CQ, Hann HW, *et al*. The management of chronic hepatitis B in Asian Americans. *Dig Dis Sci* 2011; **56**: 3143-62.

13. Han SB, Durazo FA, Saab S, *et al*. A proposed, evidence-based approach to the treatment of chronic hepatitis B. *J clin Gastroenterol* 2011; **45**: 259-66.

14. Lok ASF and Lai CL. Acute exacerbations in Chinese patients with chronic hepatitis B virus (HBV) infection: Incidence, predisposing factors and etiology. *Journal of Hepatology* 1990; **10**: 29-34.

15. Tai DI, Lin SM, Sheen IS, *et al*. Long-term outcome of hepatitis B e antigen-negative hepatitis B surface antigen carriers in relation to changes of alanine aminotransferase levels over time. *Hepatology* 2000; **49**: 1859-67.

16. Feld, JJ, Ayers M, El-Ashry D, *et al*. Hepatitis B virus DNA prediction rules for hepatitis B e antigen-negative chronic hepatitis B. *Hepatology* 2007; **46**: 1057-70.

17. Lai CL, Chien RN, Leung NW, *et al*. A one-year trial of lamivudine for chronic hepatitis B. Asia Hepatitis Lamivudine Study Group. *N Engl J Med*. 1998; **339**: 61-68.

18. Wang C, Deubner H, Shuhart MC, *et al*. High prevalence of significant fibrosis in patients with immunotolerance to chronic hepatitis B infection [abstract]. *Hepatology* 2005; **42**: 573A.

19. Nguyen MH, Trinh H, Garcia RT, *et al*. Significant histologic disease in HBV-infected patients with normal to minimally elevated ALT levels at initial evaluation [abstract] *Hepatology* 2005; **42**: 593A.

20. Lai M, Hyatt BJ, Afdhal NH, *et al*. Role of liver biopsy in patients with normal ALT and high HBV DNA [abstract]. *Hepatology* 2005; **42**: 720A.

21. Yang LM, Xu KC, Zhao YL, *et al*. Clinical significance of liver biopsy in chronic hepatitis B patient with persistently normal transaminase. *Chin J Dig Dis* 2002; **3**: 150-53.

22. Hu K-Q, Schiff ER, Kowdley KV, *et al*. Histologic evidence of active liver injury in chronic hepatitis B patients with normal range or minimally elevated alanine aminotransferase levels. *J Clin Gastroeterol* 2010; **44**: 510-16.

23. Lai M, Hyatt BJ, Nasser I, *et al*. The clinical significance of persistently normal ALT in chronic hepatitis B infection. *Journal of Hepatology* 2007; **47**: 760-67.

24. ter Borg F, ten Kate FJ, Cuypers HT, *et al*. A survey of liver pathology in needle biopsies from HBsAg and anti-HBe positive individuals. *J Clin Pathol* 2000; **53**: 541-48.

25. Kumar M, Sarin SK, Hissar S, *et al*. Virologic and histologic features of chronic hepatitis B virus-infected asymptomatic patients with persistently normal ALT. *Gastroenterology* 2008; **134**: 1376-84.

26. Keeffe EB, Dieterich DT, Han SH, *et al*. A treatment algorithm for the management of chronic hepatitis B virus infection in the United States: 2008 update. *Clin Gastroenterol Hepatol.* 2008; **6**: 1315-41.

27. Lok AS, McMahon BJ. Chronic hepatitis B: update 2009. *Hepatology* 2009; **50**: 661–62.

28. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B. *J Hepatol*. 2009; **50**: 227–42.

29. Liaw YF, Leung N, Guan R, *et al*. Asian-pacific consensus statement on the management of chronic hepatitis B: a 2005 update. *Liver Int*. 2005; **25**: 472–89.
30. Prati D, Taioli E, Zanella A, *et al*. Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann Intern Med* 2002; **137**:1-10.
31. Kim HC, Nam CM, Jee SH, *et al*. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. *BMJ* 2004: **328**: 983.
32. Seto WK, Lai CL, Philip PCI, *et al*. A large population histology study showing the lack of association between ALT elevation and significant fibrosis in chronic Hepatitis B. *PLoS ONE* 2012; **7**: e32622.
33. Ma J, Jiang Y, and Gong G. Evaluation of seven non-invasive models in staging liver fibrosis in patients with chronic hepatitis B virus infection. *European Journal of Gastroenterology & Hepatology* 2013; **25**: 428–34.
34. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperation Study Group. *Hepatology* 1994; **20**: 15-20.
35. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR cooperative study group. *Hepatology* 1996; **24**: 289-93.
36. Liaw YF, Tai DI, Chu CM, *et al*. Acute exacerbation in chronic type B hepatitis: comparison between HBeAg and antibody positive patients. *Hepatology* 1987; **7**: 20-23.
37. Perrillo RP, Campbell CC, Sanders GE, *et al*. Spontaneous clearance and reactivation of hepatitis B virus infection among male homosexuals with chronic type B hepatitis. *Ann Intern med* 1984; **100**: 43-46.
38. Lok ASF, Lai CL, Wu PC, *et al*. Spontaneous hepatitis B e antigen to antibody seroconverion and reversion in Chinese patients with chronic hepatitis B virus infection. *Gastroenterology* 1987; **92**: 1839-43.
39. Tong MJ, Sampliner RE, Govindajaran S, *et al*. Spontaneous reactivation of hepatitis B in Chinese patients with HBsAg-positive chronic active hepatitis. *Hepatology* 1987; **7**: 713-18.
40. Chu CM, Sheen IS, and Liaw YF. The etiology of acute hepatitis in Taiwan: acute hepatitis superimposed on HBsAg carrier state as the main etiology of acute hepatitis in areas with high HBsAg carrier state. *Infection* 1988; **16**: 233-37.
41. Koike K, Iino S, Kurai K, *et al*. IgM anti-HBc in anti-HBe positive chronic type B hepatitis with acute exacerbations. *Hepatology* 1987; **7**: 573-76.
42. Tassopoulos NC, Papaevangelou GJ, Sjogren MH, *et al*. Natural history of acute hepatitis B surface antigen-positive hepatitis in Greek adults. *Gastroenterology* 1987; **92**: 1844-50.
43. Iloeje UH, Yang HI, Su J, *et al*. Risk evaluation of viral load elevation and associated liver disease/cancer-in HBV (REVEAL-HBV) study group. Predicting cirrhosis risk based on the level of circulating hepatitis B viral load. *Gastroenterology* 2006; **130**: 678-86.
44. Chen CF, Lee WC, Yang HI, *et al*. Risk evaluation of viral load elevation and associated liver disease/cancer in HBV (REVEAL-HBV) study group Changes in serum levels of HBV DNA and alanine aminotransferase determine risk for hepatocellular carcinoma. *Gastroenterology* 2011; **141**: 1240-48.
45. Tong MJ, Hsu L, Chang PW, *et al*. Evaluation of current treatment recommendations for chronic hepatitis B: A 2011 update. *Journal of Gastroenterology and Hepatology* 2011; **26**: 829-35.
46. Fattovich G, Brollo L, Giustina G, *et al*. Natural history and prognostic factors for

chronic hepatitis type B. *Gut* 1991; **32**: 294-98.

47. Chen CJ, Iloeje UH, and Yang HI. Long-term outcomes in hepatitis B: The REVEAL-HBV Study. *Clin Liver Dis* 2007; **11**: 797-816.

48. Chan HL, Hui Y, Leung NW, *et al*. Risk factors for active liver disease in HBeAg-negative chronic hepatitis B virus-infected patients. *Am J Gastroenterol* 2000; **95**: 3547-51.

49. Davis GL, Hoofnagle JH, and Waggoner JG. Spontaneous reactivation of chronic hepatitis B virus infection. *Gastroenterology* 1984; **86**: 230-35.

50. McMahon BJ, Holck P, Bulkow L, *et al*. Serologic and clinical outcomes 1536 Alaska natives chronically infected with hepatitis B virus. *Ann Intern Med* 2001; **135**: 759-68.

51. Chen CJ, Yang HI, Hu J, *et al*. The REVEAL-HBV study group. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA* 2006; **295**: 65-73.

Many-to-One in Blended Learning：Imperial Learning in the Cloud Age

混合教学，集众师教一人：云端太学教学法

Ming Yang (Ph.D.)
University of Califonia, Rioverside

Introduction

1. Refletions on Imperial Learning

Imperial Learning for the prince described in the following Chinese classical passage summarizes the best Chinese imperial way in teaching a little prince to become an emperor. When the prince was a little toddler, the emperor assigned his best martial to train and protect him, his best experienced premier to be his coach, and his wisest scholar to be the teacher. The emperor created these three highest teaching positions of the country to ensure the prince to develop a strong and healthy body by learning from his lead martial, to acquire high virtues and honor from the best premier, and to learn valuable lessons from the best scholar.

As if this is not enough, the emperor created three more assistant teaching positions from his best young royal officials: best young warrior, best young experienced tutor, and best young scholar to accompany the prince during his daily life, and banquet meetings. The pupose is for the prince to learn knowledge, principles, filial piety, benevolence, propriety, righteousness conducts with best learning approach from the best men young and old. In this way, the prince is guarded against meeting bad people and shielded from seeing evil things.

《汉书》贾谊传第 26 节：昔者成王幼在襁抱之中，召公为太保，周公为太傅，太公为太师。保，保其身体；傅，傅之德意；师，道之教训：此三公之职也。于是为置三少，皆上大夫也，曰少保、少傅、少师，是与太子宴者也。故乃孩提有识，三公、三少固明孝仁礼义以道习之，逐去邪人，不使见恶行。于是皆选天下之端士孝悌博闻有道术者以卫翼之，使与太子居处出入。故太子乃生而见正事，闻正言，行正道，左右前后皆正人也。（Gu Ban）

As if this is still not enough, the emperor selected many more people from his country with all necessary knowledge and skills as assistants, helpers, and guards for the little prince，such as the people with virtuous conduct, filial piety, best learned, and approved theory. These people serve as peers surrounding the little prince. Therefore, the prince will see right things, listen to right words, grow up in right ways, because people around him are all good influences.

Teaching the prince to gain the imperial quality requires even more quality and valuable resources. From the classical passage below, Confucius, our best imperial sage and teacher, professed that formal learning for the prince should start immediately after he develops his sexual orientation. Then from the imperial School East, he learns love and care, there he understands what to love and what not to love, how and who to care as well as what to guard against. From the imperial School South, he learns trust worthiness and propriety, and there he understands what to say and what not to say, how to respect elders and how to set example for the younger ones. The prince then goes to the imperial School West, the imperial School North and finally reaches the the imperial School Center to learn the best governing theories and practices. The five imperial schools together bring the prince to his royalty and he is prepared as the best leader to govern the country.

孔子曰：少成若天性，习贯如自然。”（Confucius）及太子少长，知妃色，则入于学。学者，所学之官也。学礼曰：“帝入东学，上亲而贵仁，则亲疏有序而恩相及矣；帝入南学，上齿而贵信，则长幼有差而民不诬矣；帝入西学，上贤而贵德，则圣智在位而功不遗矣；帝

入北学，上贵而尊爵，则贵贱有等而下不隃矣；帝入太学，承师问道，退习而考于太傅，太傅罚其不则而匡其不及，则德智长而治道得矣。此五学者既成于上，则百姓黎民化辑于下矣。（Gu Ban）

After the prince graduates from the formal imperial school learning, he still needs many resources to better himself as a great emperor. From the classical passage below, we see that he is provided with many top assistants, collaborators, and educational services to maintain his highest imperial values as the future emperor ready for governing the country.

及太子既冠成人，免于保傅之严，则有记过之史，彻膳之宰，进善之旌，诽谤之木，敢谏之鼓。瞽史诵诗，工诵箴谏，大夫进谋，士传民语。习与智长，故切而不愧；化与心成，故中道若性。三代之礼：春朝朝日，秋暮夕月，所以明有敬也；春秋入学，坐国老，执酱而亲馈之，所以明有孝也；行以鸾和，步中采齐，趣中肆夏，所以明有度也；其于禽兽，见其生不食其死，闻其声不食其肉，故远庖厨，所以长恩，且明有仁也。（Gu Ban）

How can every common kid be taught today in the same best way as the little price? Traditional teaching in a traditional school obviously seems impossible to fill in the role. However, in the Cloud Age with advances of educational technology, best knowledge, skills and practices from best teachers can be selected. WIth advanced approach they can be collectively blended into the best learning methodology to serve each individual learner in the efficient way. This is what I propose below as a unique teaching methodology: “Many to one in Blended Learning.”

2. Reflections on Cloud-Based Blended Learning

With the advancement of educational technology, Net-Gen learners start to enter colleges and they brought in fresh needs and expectations to Higher Education. By carefully analyzing their needs and expectations, we propose here the project of “Scaling the Blended Learning for Net-Gen Learners”. We designed a working model of Blended Learning (BL) to reposition higher education in its fast tracks. We emphasize scaling as a way to power up the engine of blended learning so that it will shift forward the heavy load of educational paradigm.

The learner-centered learning model supported by our technology infrastructure will achieve significant increases in learning efficiency and effectiveness. This is because with this model professors can reduce their lecture preparation and presentation time from 70% to 30% to better render their lectures. So an efficiency teaching equation of 7 + 3 = 10 for F2F (Face to Face) teaching becomes a magical one: 3 + 3 = 10 for BL. Furthermore, when the 40% professor’s saved time from F2F is appropriated to a better learning for students, their success and satisfaction will increase significantly. And the BL efficiency equation will also be a magical one: 3 + 3 = 10.

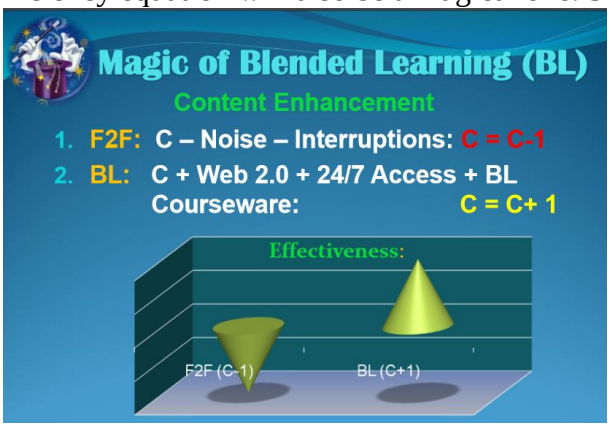


Figure 1. Content enhancement in Cloud-based Blended Learning Model

In addition to the magical increase of efficiency and effectiveness, we envision that technology innovation in BL will enhance the quality of instructional content from C = C – 1 to C = C + 1 where C stands for intended content input. With F2F teaching in a big lecture hall, part of instructional

content gets lost into noises and interruptions. So the result is expressed as C = C – 1. In BL, through content optimization, presentation concentration, and noise filtering enhancements, we have designed enhancements for the instructional content so that the formula is achieved through Content = Content + Multimedia Enhancement + Multichannel Access + Anytime/Anywhere availability. This results in an optimized content for BL that can be expressed as C = C + 1.

Cost reduction is a big advantage favoring BL. While the cost to attend a college is rising sharply and the financial support for incoming students is falling quickly, the big scissors formed by these differences will cut off many students from completing a quality education. Our BL model will significantly reduce the expected educational fees and costs from a fraction of +30% to -50%, resulting in a total of 80% reduction of educational expenses for students. The significance of reducing costs for higher education on the whole is that BL will save the public education from either becoming noncompetitive or going bankruptcy. For learners, the low BL educational costs will enable our kids from low income families to complete their quality education.

Part I: Blended Learning Framework with Updated Learning Theories

Blended learning uses a methodology to blend instructions with online delivery so that students will be able to access the content online; and they can choose the desired content modules anywhere and anytime. With blended learning, the instructional and learning paradigm shifts from instructor-centered to student-centered. The instructor more often takes a directing or coaching role, rather than the acting and teaching role. As such, educators need to harness technology and take creative approaches to address the learning needs of these Net-gen learners; understand their new learning behavior; and provide assistance and solutions around their preferences.

Instructors should also understand that while technology and online delivery free them from these routine tasks of preparing, teaching, and testing, they should spend more time creating individualized intervention plans, increasing the timeliness and frequency of direct instruction for the students who need it most.

Without specialized training in blended learning course design and methodology, instructors may not take advantage of technology and online delivery in their traditional approach. Furthermore, without spending more time helping students individually, Blended learning processes only make teaching easy and convenient for the instructor, without increasing effectiveness in student-centered learning.

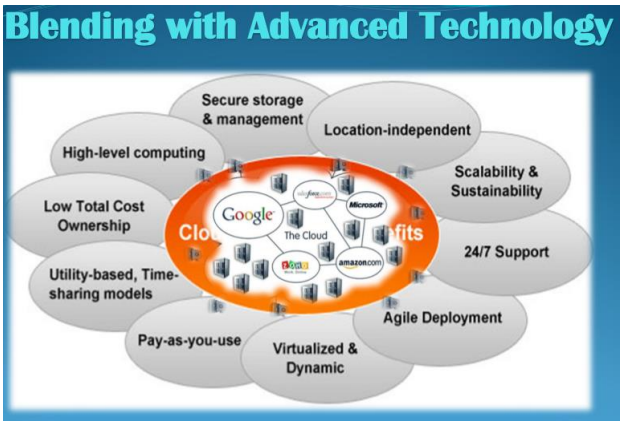


Figure 2. Advantages of Cloud-based Blended Learning Model

With the advancement of educational technology, Net-Gen learners start to enter colleges and they bring new challenges and expectations to Higher Education. Higher education in the technology era can not afford to ignore the social and economic changes driven forward by technology. The teaching paradigm has to embrace technology advancement so as to satisfy expectations of net-gen learners. As technology doubles its advancement almost every two years, education should not continue its way of teaching as it had been without it for 150 years.

Our vision is that teaching paradigm is bound to be shifted by technology advancement as well as the net-gen college learners. The power and energy of technology has been so vast and far-reaching to net-gen learners that higher education should be repositioned to make a sea change to embrace them. The new paradigm shifted by blended learning will be one that is student-centered, community-embraced, and with finger-tip access of content. The enclosed classroom becomes open in 24/7 in the new learning environment. Quality education is no longer enclosed within a campus but available anytime and anywhere. Students can reach their destination through their preferred paths, not just one pre-defined path that requires students to follow, with success or failure, as described in the following 4 dilemmas.

Dilemma 1: Learning at a Shifted Teaching Paradigm

Blended learning is a technology enhanced learning framework that shifts conventional teaching to a learner-centered approach by mixing class instructions and learning activities with multimedia presentation and online multichannel delivery. Students in a blended learning program will be able to access course instructions and content at their finger tips, choose the desired content modules anytime, and interact with peers and the instructor anywhere.

With blended learning, the instructional and learning paradigm shifts from instructor-centered, class-based, and textbook-dependent framework to a student-centered, community-based, and media dependent framework. In a blended course, the instructor often takes a role of a coach, rather than that of a presenter. As such, educators need to be savvy in technology and develop creative approaches to address learning needs of the Net-gen learners; understand the new learner-centered learning behavior; and provide assistance and solutions around their preferences.

However, instructors also need to understand that, they should be prepared that while multimedia technology and online delivery can free 30-70% of their time from routine tasks in preparing, teaching, and testing their classes, they will also need to spend more time creating individualized intervention plans, lead class discussions and group interactions, and increase the timeliness and frequency of specialized instructions for the students who need them most.

Without training instructors in appropriate course design and methodology in blended learning, the purpose of blended learning may not be well served in that instructors may simply apply technology and online delivery to their traditional teaching methods. Without spending more time with students individually, technology can only make the instructors' teaching easier and more convenient at most, without substantially increasing learning effectiveness in the students' learning process.

Recent findings show an increasing trend in online learning and a declining trend in faceto-face learning gradually moves away from each other. Furthermore, online learning is gaining momentum. For example, companies similar to the University of Phoenix (UOP, under Apollo Group) increase in value, while the poor economy has reduced the value of most other companies. This situation may have resulted from the University of Phoenix increasing their online programs and the access of higher education online became a much favored approach for students who had to work while earning a degree.

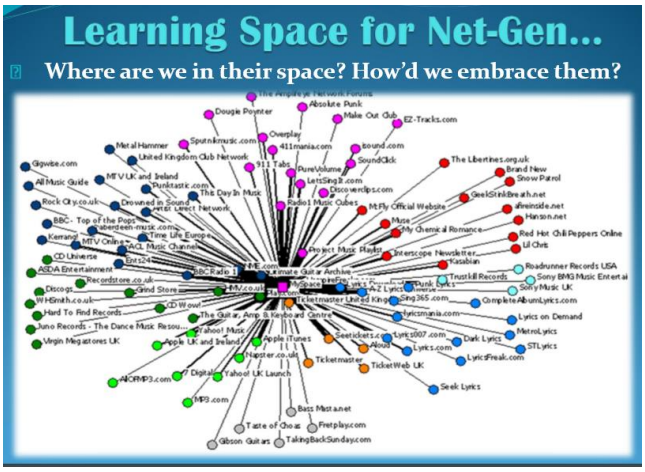


Figure 3. Simulation of learning space for Net-Generation in Cloud-based learning

The teaching paradigm also shifts with wide acceptance of new technologies on campuses. Education has been influenced by continued expansion of personal learning devices such as laptops, smart phones, iPods, and iPads among learners. Institutions have adopted course management systems for academic content, Internet networking tools, collaboration tools, and community building tools for interactivity. Here are some key technology drivers that influence changes in education: Ubiquity of Internet access, gig-byte connectivity, and Web 2.0 functionality. Advances in graphics, multimedia and interactivity: authoring, publishing, lecture capturing, and content delivery in multiple modes and channels.

Dilemma 2: Large Courses with Less Individual Attention

Either an online or a face-to-face learning model alone may not maximize student success in today's learning environment. With classroom teaching serving as the norm in colleges and universities, online teaching is often undervalued and mostly used to supplement or fill the void of direct teaching. On the other hand, classroom teaching itself has limitations. The most notable factor is that classes become much larger. Faculty members and students feel that teaching, discussion, and interactions become unmanageable and ineffective.

This circumstance causes students to watch presentations passively, with little interaction, Instructor do not have enough time to give enough direct guidance and individualized attention to students.

Let's take UC Riverside as an example. In general, undergraduate classes at University of California Riverside are generally big. Each quarter there are about 200 courses with enrollment over 100 students.

Quarter	Courses over 100 Enrollments	Total # Enrollments	Enrollments per Course
Winter 2010	196	46,000	235
Spring 2010	186	45,000	242
Fall 2009	216	49,000	227

Table 1. Undergraduate currilum full of large classes in a typical university

Let's take a UCR first-year business student, Cole Brian, as an example. Out of 10 courses he has taken so far, six out of 10 had a class size of more than 200 students, averaging 230 students per class. When each course accounts for about 1/3 of an instructor's total work time in a quarter, how much time would Cole Brian be allowed to speak to his instructors? Well, here is the data:

BUS_100_001_09F	335	BUS_100_022_09F	59
BUS_101_001_09F	342	BUS_101_023_09F	61
BUS_103_001_09F	346	BUS_103_023_09F	59
BUS_106_001_10W	262	BUS_104_122_10U	57
BUS_109_001_10S	278	BUS_106_022_10W	12
BUS_112_001_10S	87	BUS_109_022_10S	69
CS_008_002_10W	169	CS_008_024_10W	35
ECON_103_001_10W	346	ECON_103_023_10W	39

Table 2. Large classes in a undergraduate currilum

With such big classes, one can reckon why C instructor may not be able to help him individually for a half hour during the whole quarter.

Cole's TA sections are also alarmingly large, averaging 45 students per discussion section. Maybe a TA can afford to help him individually for one hour during the whole quarter.

BUS_100_022_09F	59
BUS_101_023_09F	61
BUS_103_023_09F	59
BUS_104_122_10U	57
BUS_106_022_10W	12
BUS_109_022_10S	69
CS_008_024_10W	35
ECON_103_023_10W	39
MATH_004_002_10W	32
MATH_022_014_10S	34

Figure 3. Discussion sections are too large in a ndergraduate currilum

For students in this dilemma, blended learning offers a solution. When class teaching is supplemented with online lecture captures, communication tools, and interactive functions, a student's access to course content becomes frequent and instant. With a virtual community where a student's classmates, peers, group members, TAs and the instructor are instantly available, the student's learning process becomes authentic, manageable, and effective.

Time and space can be simulated to make teaching an amazing art. For example, if Albert Einstein's lectures had been captured, students would not have to depend upon anecdotes, stories, and readings to learn from him. They would have learned with the great master by simply repeating the lectures as comfortably, as conveniently, and as frequently as they would like.

Dilemma 3: Learning at Higher Costs and Less Resources

Low-income young adults who need to combine work with learning may benefit substantially from blended learning programs. Under the pressure of the economic downturn, college students face the dilemma of learning at a record level of expenses which further pressures their personal lives. Recent headline news at the LA Times revealed that some UCLA students became homeless. They could not afford rent for a dorm and had to sleep in their cars at night. They had to either skip meals or depend upon donated meals to stay out of hunger.

Let's take an UCR MBA student, Jayanti Dash, as an example. She came from India last year to attend the UCR MBA program. She was married and her loved ones were left behind in India. In 2009, the annual fees and costs for an international MBA student at UCR amounted to about $48,000. And tuition is still on a sharp rise. She had to spend all savings from her family; ask help from her uncle's family in the US; and work part time at UCR to make ends meet. Jayanti would have to quit

her studies if she failed to compete and win the International scholarship from her business school, AGSM.

Let's see what it means for her if AGSM had offered MBA courses online with authentic lecture captures through our course management system. She could view the same lectures by the same instructors as others who attended the classes. With the virtual classroom and video conferencing, she could meet online and interact with the professors and other fellow MBA students, and complete the courses without leaving her home country. She would complete the MBA program while spending much less money, getting more individualized instruction, staying home with loved ones, keeping herself free from oversea anxieties and barriers, and excelling with a better learning outcome in the end. Isn't this a better way for education to help fulfill a student's dream and shape a person's life?

Students learn better when they are motivated, inspired, and individually taught. Their learning outcome is better when they have abundant learning opportunities and redundant resources. Learning is more fun and effective when it becomes student centered. These claims are supported by research, experiments, psychological testing, and learner surveys. Here, I want to point out that our blended learning model is deeply rooted in these mature learning theories. The benefits that the model offers to students go beyond just the learning outcome. The model helps the low income students reach their dreams.

Dilemma 4: More New Challenges against Quality Education

While more universities are considering the commencement of online programs, some faculty are offended and complain that online teaching is moving away from their traditional teaching and leaving them behind. A notable example is that, quite recently, Christopher Edley, dean of Berkley law school, proposed a Berkley online program, which immediately incurred criticisms from faculty as a degradation in education. This response is expected since currently education is defined and decided mostly by educators, not by learners, or by technology innovators.

However, the paradigm shift is obvious and it really does not take an evolution to bring some educational programs online. With the Net-gen learners reaching traditional college and university age, they will be the tipping point in favor of blended learning. For them, libraries can be online; lectures can be fully captured and accessed online; and class activities can be conducted online; as all students are already online. So it makes perfect sense for them to take their courses and complete their educational programs online.

Additionally, aligning the finances to facilitate the implementation of effective blended learning may require new thinking about course scheduling, faculty compensation, institutional budgeting and revenue-sharing. Online courses have to be scheduled and completed with high academic standards. Intellectual property needs to be honored even when it is online. Faculty work and effort have to be duly compensated. Online programs need to be in line with the university budget. Only when online programs are wellintegrated into the university's infrastructure, will they gradually improve in quality and recognition.

However, making online programs widely available on a cost-effective basis with a high level of consistency remains an obstacle. Recently, several online programs have failed at some high profile universities. The failures were not necessarily caused by lack of insight or desire to start online programs but by poor design and an inappropriate approach. There are three lessons to learn: First, using online programs to just drastically reduce the number of instructors and educational budget is bound to fail. Second, degrading education by putting low quality courses or deserted courses online is another major factor in the failure of online education. Third, courses posted on line without involvement of the instructors may cause the course content to suffer. Finally, poorly designed online courses with insufficient or ineffective technology would add barriers between instructors and students so as to block them from success.

Part II: Designing Principles and Methodology in Blended Learning

Our model will integrate a wide range of instructional functions into a technology enabled framework. It is designed to empower learners with self-controlled content access, community based participation, individualized interaction, as well as up to 70-30 percent ratio in informal vs. formal learning activities. The informal learning activities will guide learners to "navigate from informational to instructional content, from skills assessment to supportive tools, and from coaching to collaborative environments."

Our pioneering belief is that blended learning will be the mainstream in higher education for the Net-Gen learners. We believe that magic will happen in a blended learning environment when we serve authentic lecture captures, open content sources, and multimedia input with redundancy and multiplicity. This approach streamlines upscale technology in blended learning in ways that have been shown to propel learning, not to inhibit it. It uses technology to orchestrate learning functions and virtual communities so that students are less dependent on physical participation and campus resources.

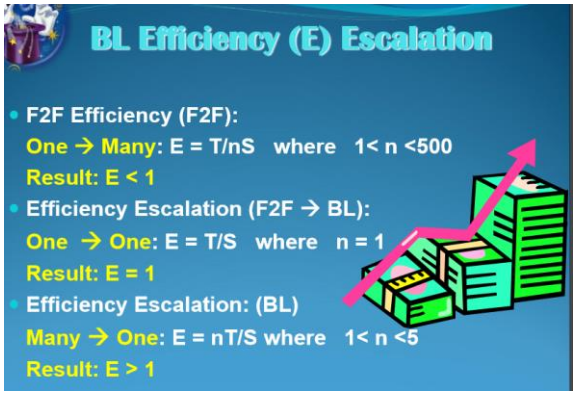


Figure 4. Efficiency escalation in Cloud-based Blended Learning Model

We propose to extend online network and connectivity in blended learning to cultivate virtual participation, collaboration, peer support, and other social learning styles. We intend to design learner-centered and performance-based courseware to encompass individualized learning needs, and combine formal and informal learning. The following five design principles of blended learning for "net generation" learners are deemed most effective in the student learning success.

1. **Make magic happen in blended learning where authentic lecture captures, open content sources, and multimedia input are served with redundancy and multiplicity.**

Teaching is where magic happens. Let's take Albert Einstein as an example both as a teacher and a learner. We used to hear the story about Albert Einstein's driver being able to deliver his lecture flawlessly. But the driver had to refer to a disguised Einstein to answer a student's question. Now, let's replace Albert Einstein's driver with authentic lecture captures. I assure you that lecture captures can deliver the lecture better than the driver, because every bit of the capture matches the authentic lecture.

This would enable the professor to be freed from mundane lecture constraints and to interact more effectively with the audience. Imagine if Albert Einstein's lecture had been captured and played to our students, and students saw Einstein sitting at the coffee table, putting a biscuit on the right, a bagel in the middle and then his coffee cup to the left. He then drew a line with his finger in between. Voila, the famous atomic energy formula came into being. The atomic era started thereafter. If Einstein's greatest ideas could be replayed and illustrated in the lecture capture, don't you think that it would capture the brilliance of his lecture, inspire the students, and spark their minds?

Now let's capture Albert Einstein's learning as a student. He was not brilliant in his early years. He took tests and tried to get into an elite college (somewhat comparable to Stanford University), but he failed. Instead, he was admitted to a less elite college. However, he was fortunate to have a professor there who spent a lot individual time with him. Einstein was inspired by him; became more confident; and acquired the mindset of a scientist. There Einstein shined brilliantly, and a new star was formed.

What is the moral of this story to our current professors and students? For professors, their lectures are the great assets of their wisdom and knowledge. Some of them have worked their whole life to collect these assets. Don't they want them to be saved and continued to educate students in the future? When great masters of knowledge pass away, they fade into history, like stars falling from the sky.

However, with captured lectures, this reservoir of instructional content with enhanced media will surely be a legacy for the instructor and will also contribute to the intellectual wealth of the university. For students, course captures served through a learning management system enables an instructor's lectures to be reviewed, re-deployed, transferred and re-used. With course captures, students will have access to a huge wealth of intellectual assets that are collected, built upon, and refined by great teachers. When students can follow and learn directly from great masters of their time, there will be chances that the students can be brought up as masters of knowledge in the future.

2. Upgrade upscale technology in blended learning used in ways that have been shown to propel learning, not inhibit it.

There are a few situations where technology may not help learning. First, if the learning costs are increased due to technology intervention, student from low income families will suffer. Also, when the financial resources of a learning institution are scarce or bleak, it does not make sense and is not the right time to switch to innovative methods like blended teaching.

Second, political determinants often come into play and certain technology may not be allowed freely in some countries. For example, Internet usage, social networking, as well as freedom of opinion are tightly checked in China. Blended learning will not work if free online access is not granted. Blended learning may not be effective when political heavy hands intervene.

Third, learners' resistance to innovation must also be considered if a certain technology creates a steep learning curve. Oftentimes teachers oppose technology in their courses because they are not sure how to use it.

Fourth, learning styles may also be affected by technology. Without proper consideration of how the student is going to learn, an implementation of technology may adversely affect a student's learning ability.

Figure 5. Efficienct content preparaton in Cloud-based Blended Learning

Finally, other factors such as student demography, institutional and organizational structures, human resources, and communication should be also considered to ensure that technology is scaled properly to promote learning. The learner-centered learning model supported by our technology infrastructure will achieve significant increases in learning efficiency and effectiveness. This is because with this model professors can reduce their lecture preparation and presentation time from 70% to 30% to better render their lectures. So an efficiency teaching equation of 7 + 3 = 10 for F2F (Face to Face) teaching becomes a magical one: 3 + 3 = 10 for BL.

Furthermore, when the 40% professor's saved time from F2F is appropriated to a better learning for students, their success and satisfaction will increase significantly. And the BL efficiency equation will also be a magical one: 3 + 3 = 10.

3. Design an effective framework in blended learning to decrease costs

Blended Learning Framework orchestrates learning functions and virtual communities so that students are less dependent on physical participation and campus resources. The technology-based framework will be one that is system-based, supportive, transparent, and easy to use. It should not create barriers or learning curve which may increase learning burden and impede the success of technically- disadvantaged students.

If students are learning and networking on their computers, they should not be required to attend a class in order to pass it. Technology should be used to extend classroom teaching into homes, offices, public areas, and the general community. Technology should be used to extend the learner's access to course content, interaction, participation, and discussion at any time during the day and the week. When all these opportunities are provided to students as options to reduce their learning cost, many students will take advantage of it. When students find that being technology savvy really benefits their learning and contributes to their success, blended learning will enter a new era and will become the norm for the net gen learners. We are not just hoping for it, I see this era is indeed coming.

Figure 6. Learner cost reduction in Cloud-based Blended Learning

Cost reduction is a big advantage favoring BL. While the cost to attend a college is rising sharply and the financial support for incoming students is falling quickly, the big scissors formed by these differences will cut off many students from completing a quality education. Our BL model will significantly reduce the expected educational fees and costs from a fraction of +30% to -50%, resulting in a total of 80% reduction of educational expenses for students.

The significance of reducing costs for higher education on the whole is that BL will save the public education from either becoming non-competitive or going bankruptcy. For learners, the low BL educational costs will enable our kids from low income families to complete their quality education.

4. Extend connectivity in blended learning via online network to cultivate virtual participation, collaboration, peer support, and other social learning styles.

When a virtual learning environment is created, expenses to attend classes and use other shared resources can be greatly reduced. The significance of blended learning means a lot more to students than to instructors. While a learning management system such as Blackboard Learn System is a great system for instructors to present and deliver coursework, a lecture capturing system will serve as a scaffold for students to reach higher, access more, learn more effectively, and achieve a better learning outcome. This is especially true when students now endure large class sizes with separate sections taught by teaching assistants.

The adverse learning factors for low income learners include time conflict, scarce resources, lack of access, and low incentives. While being learner-centered can help them to resolve time conflict,

and motivate them to learn better, a home-based learning system will provide them with more resources and better access to content, with facilitated by technology.

5. Design learner centered and performance based courseware to encompass individualized learning needs in blended learning.

When learning is instructor-centered, all learners have to come to the instructor and compete for his time in the lecture and discussions. When learning is in reception mode, students' learning mostly begins after the teacher is done with his lecture. Many instructors consider that their teaching is completed when the lecture is presented. Actually, instruction should be present while the learning in happening. This is why individual attention given to students by the instructor is vital in the student's learning process.

In blended learning with technology, authentic lectures can be captured and replayed in multimedia and multiple channels while the learning is happening. An instructor will be able to extend his or her help to students in a way that only individual attention can achieve. Authentic lecture captures do not just a replace or supplement class presentations. They help instructors in three ways to maintain higher standards in their lecture presentations.

- First, instructors who use lecture captures would have strong consciousness of self image. Thus they should appear better before students.
- Second, instructors can review the lecture captures to see where they do best and where they should improve. They become more careful with the content and the time to present it. So the lectures are captured with positive result. And with more practice over time the lectures would be improved dramatically. By the way, all the capturing technology lies in the background so no one can notice it or be interfered by it.
- Finally, the instructor has the ultimate control over whether to post the lecture captures to students or not.

Unless the instructor wants to take the lecture capture to the department administrators for positive uses, administrators would generally have to obtain the instructor's permission to go into the course to view the captures.

Conclusion：

The use of cloud-based Blended Learning technology to improve the learner experience at points of transition is reflective of wider research that identifies key components of effective transition and approaches. This includes pre-entry support, social and academic integration and skills development, timely information provision, student support and offering students choice and control/empowerment (Whittaker).

The intended outcomes of this Many-to-One Blended Learning model are very significant and far reaching. We will excel and realize all 5 outcomes outlined by the NGLC foundations. This project will establish as an entity a BL R&D infrastructure in the form of a cloud-based institute in an 8-year cycle with four phases:

1. Establishing a working BL Framework by proving concept, developing, testing, and evaluating its intended outcomes at 5 selected universities or programs;
2. Setting up BL technology infrastructure and system network including program expansion at designated colleges and universities, establishing top level technology grid with servers and managing systems, establishing a formal BL model and framework, and setting up a BL center of research;
3. Centralizing the BL infrastructure and its programs, including building up basic infrastructure for the BL Institute, starting accreditation of BL programs, and sponsoring selected BL programs at some designated learning institutions;
4. Completing cloud-based Institute of Blended Learning by establishing learning centers and running BL programs at national and global level. (NGLC 2013

References

1.Ban, Gu, (80 AD) 《汉书》贾谊传，*Biography of Yi Jia*,"*History of Han Danasty*" available from http://ctext.org/han-shu/jia-yi-zhuan/zhs, 2010-2016

2.Confucius（722-481 BC）《孔子集语》卷三, 孔子家语, available from http://www.hbrc.com/news/key_149580_2.html，2014

3. Yang, Ming, (2013) *Scaling the Blended Learning for Net-Gen Learners,"*Technology grant proposal for "Next Generation Learning Challenges"

4. Lefever， R. & Currant， B (2010) HOW CAN TECHNOLOGY BE USED TO IMPROVE THE LEARNER EXPERIENCE AT POINTS OF TRANSITION? available from http://technologyenhancedlearning.net/files/2010/04/ELESIGliteraturereviewFINAL240210.pdf

5. Whittaker, R. (2008) Quality Enhancement Themes: The First Year Experience, Transitions to and during the first year, Quality Assurance Agency for Higher Education, available from: http://www.enhancementthemes.ac.uk/documents/firstyear/Transition%20-%20Final.pdf

6. NGLC, (2013) Next Generation Learning: The Pathway to Possibility, available from http://net.educause.edu/ir/library/pdf/ngw1301.pdf

Microstructure based evaluation of material fracture toughness

Yan Li*

**Department of Mechanical and Aerospace Engineering, California State University, Long Beach, CA 90840, USA*

Summary A cohesive finite element method (CFEM) based multi-scale framework for analyzing the effects of microstructural heterogeneity, phase morphology, constituent behavior and interfacial bonding strength on the fracture toughness of materials is developed. The computational framework allows the fracture toughness to be predicted as functions of microstructural attributes and constituent behavior. The method uses the *J*-integral and the linear elastic relation between *J* and *K*.

INTRODUCTION

The evaluation of material fracture toughness as functions of microstructure attributes is a fundamental issue in material science. It requires proper characterization of microstructures and quantitative correlation of microstructural attributes with overall material response. Experimental approaches allow systematic studies of microstructure-fracture toughness relations but fail to explore material configurations not yet in existence. Besides, the trial-and-error process is usually associated with high cost and long cycles. To overcome the above shortcomings, a cohesive finite element method (CFEM) based multi-scale framework for analysing the effect of microstructural heterogeneity, phase morphology, constituent behavior and interfacial bonding strength on the fracture toughness is developed as shown in Fig. 1. This framework allows explicit representation of microstructures and allows explicit representation of microstructures and account of microstructural level of deformation, damage and failure mechanisms, while allowing macroscopic conditions such as controlled loading and structural response to be considered at the same time. This framework provides a means for calibrating model parameters at the microscale through macroscale responses which can be easily measured in experiments. More importantly, it allows macroscopic response measures such as K_{IC} to be evaluated as functions of microstructure by calculating the *J*-integral along an arbitrary contour within the homogenized region. Although the framework developed here can be applied to any material system in principle, computations in this paper concern Al_2O_3/TiB_2 ceramic composites. These two-phase materials consist of an Al_2O_3 matrix and a TiB_2 reinforcement phase. Microstructures with more than two phases can also be analyzed. Based on the numerical results, relations between microstrucutral parameters and fracture toughness are established. These relations can be used for the selection of materials and the design of new materials with tailored properties.

J-INTEGRAL BASED FRACTURE TOUGHNESS EVALUATION

For the model shown in Fig. 1, contours for *J* evaluation are solely within the homogenized part of the specimen where no cohesive elements are used. For brittle materials, the *J*-integral is equivalent to the energy release rate *G* and can be related to the stress intensity factor *K* [1]. The generalized path-independent *J*-integral for dynamic conditions is used in this analysis as [2–4]

$$J = \int_{\Gamma}\left[\left(\int_0^t \boldsymbol{\sigma} : \mathrm{d}\boldsymbol{\varepsilon} + \frac{1}{2}\rho\dot{\mathbf{u}}\cdot\dot{\mathbf{u}}\right)\mathrm{d}x_2 - \boldsymbol{t}\cdot\frac{\partial \mathbf{u}}{\partial x_1}\mathrm{d}s\right] + \int_A\left(\rho\ddot{\mathbf{u}}\cdot\frac{\partial \mathbf{u}}{\partial x_1} - \rho\dot{\mathbf{u}}\cdot\frac{\partial \dot{\mathbf{u}}}{\partial x_1}\right)\mathrm{d}A, \qquad (1)$$

where ***t*** is the traction on a surface with normal **N** in the reference configuration, **u** is the displacement,
ε denotes the engineering strain and is the mass density.

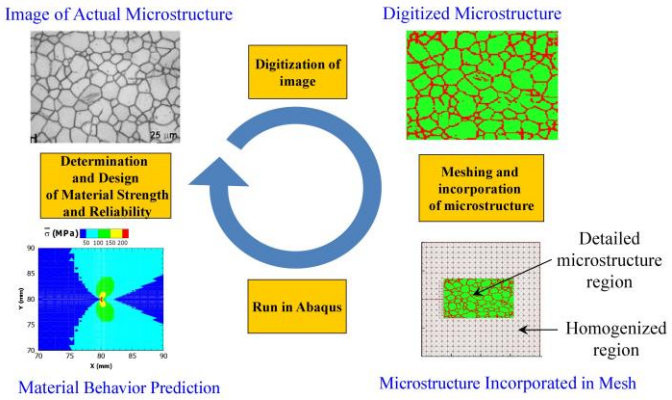


Fig. 1 The multi-scale CFEM framework for material design.

Fig. 2 shows six snap shots of the crack propagation process in a microstructure with circular TiB2 reinforcement at a loading velocity of v 5 mm/s . The corresponding histories of *J* and *K* are shown in the center. Fracture initiates in the Al2O3 matrix at 105.0 μs [Fig. 2(a)], this event defines the initiation toughness K^i . The crack is arrested by a TiB2 particle and pauses at the Al2O3/TiB2 interface
for approximately 42.5 μs [Fig. 2(b)]. During the pause, *J* increases rapidly. At approximately 149.2 μs [Fig. 2(c)], as a result of the higher level of driving force *J*, the crack penetrates the TiB2 particle. Subsequently, the crack propagates rapidly, causing *J* (and therefore *K*) to plateau for the remainder of the analysis. The average value of *K* during this period is taken as the propagation toughness K_{IC} [5, 6].

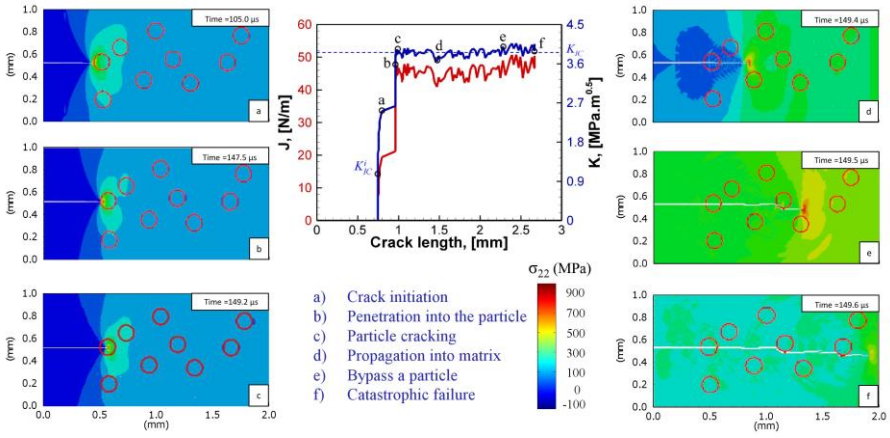


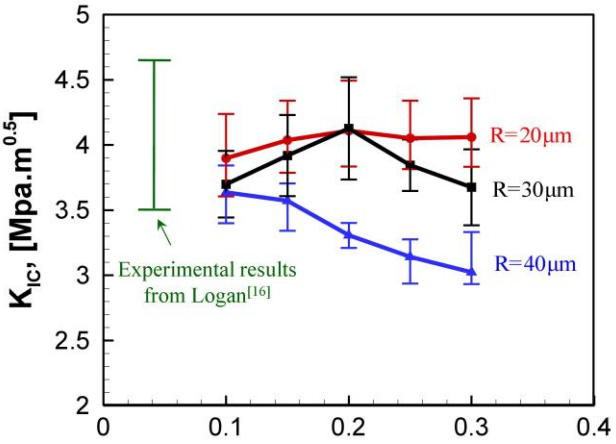
Fig. 2 Crack propagation in a microstructure with circular reinforcement and the evolution of *J* and

K during this process.

RESULTS AND DISCUSSION

Calculations carried out concern Al2O3/TiB2 two-phase ceramic composites and focus on the fundamental fracture mechanisms during crack initiation and propagation. Results of CFEM calculations show that both microstructure and constituent properties can significantly influence fracture behavior and combine to determine the overall fracture toughness through the activation of different fracture. As shown in Fig. 3, microstructure affects K_{IC} much more than K_{IC}. The finest reinforcements give rise to the highest propagation toughness, but have the least influence on the initiation toughness. The opposite trends in influence have to do with how cracks interact with particles. If a large particle happens to be at the tip of the pre-crack, a higher level of stress is required to initiate the crack and, consequently, the initiation toughness is higher. Such events are less frequent, as shown in Fig. 3. On the other hand, a propagating crack is more likely to penetrate a large particle, causing immediate catastrophic failure of the material and limited improvement of the propagation toughness.

The finest



Particle Volume Fraction

Fig. 3 Propagation toughness as a function of reinforcement size and volume fraction. The error bars indicate scatter of results obtained from twenty random instantiations (samples) of each microstructure.

The CFEM results in Fig. 4 show that the strongly bonded (stiff) interface ($Q = 10$) leads to extensive particle cracking and a 14.7% decrease in K_{IC} relative to the baseline case ($Q = 1$). As the bonding strength decreases (or interfacial compliance increases), more crack deflection into the matrix/reinforcement interfaces occurs, causing K_{IC} to increase initially, but ultimately decrease between $Q = 10^{-3}$ and 10^{-5}. The highest K_{IC} occurs at approximately $Q = 10^{-3}$ which corresponds to $T^{in}_{max} = 0.6$ and $\Delta_{tc} = \Delta_{nc} = 0.068$ μm. Particle cracking can be effectively avoided when the interfaces are weak or quite compliant ($Q < 1$). As Q decreases, interface debonding gradually outweighs matrix cracking and becomes the dominant fracture mode. However, excessively compliant interfaces ($Q \leq 10^{-5}$) provide easy pathways for crack growth and lead to lower K_{IC} values. This scenario is consistent with what has been reported for porous ceramic materials whose fracture toughness values are lower than two-phase or even single phase ceramics.

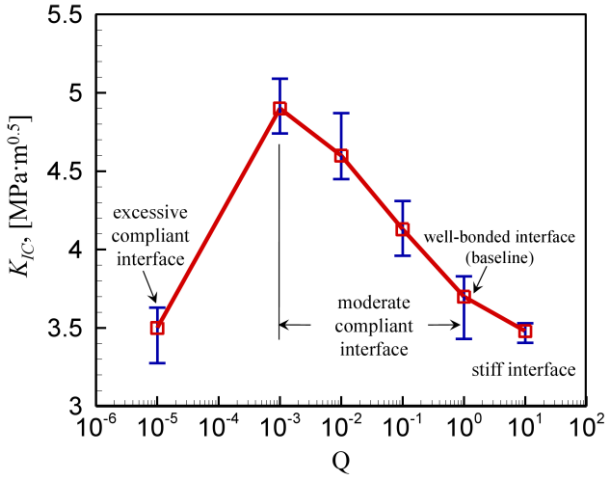


Fig. 4 Effect of the compliance or bonding strength of the reinforcement-matrix interface as measured by the strength ratio Q on K_{IC}. The error bars indicate scatter of results obtained from twenty random instantiations (samples) of each microstructure.

Interface debonding is the most beneficial fracture mechanism and is primarily promoted by small reinforcement size, rounded particle shapes and appropriately bonded and compliant einforcementmatrix interfaces. In contrast, particle cracking, which triggers catastrophic material failure, usually occurs in microstructure with large reinforcement particles, lower particle roundness and overbonded/stiff interfaces. Important constituent parameters are the fracture toughness of the matrix phase and the toughness of the interface between the matrix and the reinforcement phases.

CONCLUSIONS

CFEM calculations show that microstructure and constituent properties can significantly influence fracture behavior. Microstructure size scale refinement, attainment of more rounded reinforcement morphology and appropriately balanced interfacial bonding strength can lead to enhanced material fracture toughness. The CFEM framework is useful tools for the design of failure-resistant materials.

References

[1] T. L. Anderson, *Fracture mechanics: fundamentals and applications*. CRC Press, 1995. [2] B. Moran and C. F. Shih, "A general treatment of crack tip contour integrals," *Int. J. Fracture.*, vol. 35, no. 4, pp. 295-310, Dec. 1987.

[3] L. B. Freund, *Dynamic fracture mechanics*. Cambridge University Press, 1998.

[4] M. Zhou, G. Ravichandran and A. J. Rosakis, "Dynamically Propagating Shear Bands in Impact-loading Prenotched plates, II - Numerical Simulations", *J. Mech. Physics. Solids*, 44, pp. 1007-1032, 1996

[5] Y. Li and M. Zhou, "Prediction of fracture toughness of ceramic composites as function of microstructure: I. Numerical simulations," *Journal of the Mechanics and Physics of Solids,* vol. 61, pp. 472-488, 2// 2013.

[6] Y. Li and M. Zhou, "Prediction of fracturess toughness of ceramic composites as function of microstructure: II. analytical model," *Journal of the Mechanics and Physics of Solids,* vol. 61, pp. 489-503, 2// 2013.

实现人力资本增值的最大化
—企业人力资本配置实证研究

刘绛昊

江苏省丹阳市政府法制办公室，江苏 丹阳 212310

摘要：实现增值最大化是人力资本研究的一个重要课题。本文认为，企业人力资本的增值主要取决于优化配置。基于2008—2012年在500家企业的调查数据，本文运用《周易》阴阳五行生克关系原理，以人力资本绩效考评为中心线索，对企业各类人力资本配置现状进行了测度并计量分析，分别建立了核心、关键和一般人力资本的最佳配置模型。研究表明，配置对人力资本实现增值最大化具有决定性作用，是企业实现收益最大化的主要原因和重要路径。

关键词：人力资本配置；实证分析；模型

To maximize the value of human capital
- An Empirical Study of Human Capital Configuration

Abstract: Achieving value maximization of human capital is an important research topic. This paper argues that the value of enterprise human capital depends on the configuration. Based on the survey data in 500 companies during 2008 to 2012, the paper uses the *yijing*(Book of Changes) principles of *yin yang* and the mutual supplementing or destroying relationship between *five-element* , with the human capital performance evaluation, to give a measurement and quantitative analysis onthe configuration status of the various categories of human capital .A optimalconfiguration model is established between three kinds of human capital : core, critical and general. This research manifests that proper configuration to maximize the value of human capital can play a decisive role while being the main reason andan important way for companies to maximize revenue.

Keywords: Human Capital Allocation Empirical Analysis Model

一、问题的提出

资本是用来增值的，人力资本也是如此。对企业而言，引进什么样的人力资本可以增值？采取什么方法能使人力资本实现增值最大化？这是企业人力资本运营中的一个重要内容。纵观国内外的理论文献，对企业人力资本的增值机制和方法关注较少；增值的主要原因被笼统归结到教育和培训等外部形态；一些增值模型缺乏应用性，让企业难以接受和操作。因此，从1995年开始，本文着手对企业人力资本如何增值进行了长达20年的研究。通过研究发现，企业对人力资本进行优化配置，是实现增值最大化的关键因素。

所谓人力资本配置，是指企业为了使人力资本投入达到产出和收益最大化，按照人力资本与外在环境的作用关系，对引进的人力资本进行有计划、不定期、最优化地配备和布置。配置是企业一种自觉的主体行为，具有主动性、计划性、控制性和动态性的一般特征。外在环境由“人”和“物”两类因素组成，本文专指“人”的因素。

本文通过对人力资本在运营过程中与外在环境的作用关系进行计量分析，揭示配置对人力资本实现增值的决定性作用，探索建立一个实用性强的人力资本最佳配置模型。

二、中国古典文化的启示和运用

人力资本实现增值最大化，关键取决于优化配置。这一发现，主要来自《周易》文化的启示和运用。西方文明有《圣经》，东方文明有《周易》。《周易》是中国古代最早的一部哲学巨著，是中国古典哲学的源头，被誉为中华民族文化的智慧宝典。由于它述及了许多经济方面的内容，又堪称是中国乃至世界上最古老的一本经济学著作。《周易》对经济学方面有着许多重要论述，限于篇幅，不便一一赘述。值得本文关注的,主要有两点：

（一）阴阳五行学说

阴阳是中国古代哲学的一对范畴。国内大量研究文献表明，五行是由阴阳相互作用而生化出来的一种实质性的事物形态，用金、木、水、火、土五类最基本的功能属性来标示和归纳，体现了阴阳的不同组合状态，标志着阴阳相互运动发展变化的规律。①[1]阴阳五行学说，是《周易》文化体系中的重要基石。

《周易》指出，在阴阳五行运动中，对立统一的双方存在着两种基本关系：相生或相克。这种相生或相克关系，都是主体对客体的单向作用关系。所谓相生，是指主体能够为客体提供全面的关心、支持和帮助，使客体始终处于上升或前进趋势，呈现出主体对客体的友好关系；所谓相克，是指主体对客体进行全面排斥、制约和打击，使客体始终处于下降或后退趋势，呈现出主体对客体的斗争关系。

五行之间相生是，水生木，木生火，火生土，土生金，金生水；五行之间相克是，水克火，火克金，金克木，木克土，土克水。在相互作用关系下，双方力量就会发生一定的变化，要么增强，要么减少，从而引起一切事物的运动、变化和发展。例如，在水与木的关系中，水生木，木的力量在水的帮助下增强了（相生）；在水与火的关系中，水克火，火的力量在水的打击下减少了（相克）。根据五行生克关系之间相互作用的规律，能够推演出各类事物或现象的复杂联系及运动变化的结果。

（二）六十甲子纳音表

六十甲子纳音表,是阴阳五行学说的最高表现形式，也是《周易》理论体系中的重要载体之一。中国古代从黄帝时期开始，就用天干、地支来计时，形成了中国独有的农历，至今仍在使用。所谓十天干，是指甲、乙、丙、丁、戊、己、庚、辛、壬、癸，十二地支，是指子、丑、寅、卯、辰、巳、午、未、申、酉、戌、亥，把它们按照两两相配的原则，以一个天干和一个地支相配合，天干在上，地支在下，用来计年、计月、计日、计时。由于天干由甲起，地支由子起，阳干配阳支，阴干配阴支，从甲子到癸亥一个循环下来，共有六十个组合，刚好是 60 年，称之为六十甲子。所谓纳音，就是把中国古典声律中的五音（宫、商、角、徵、羽）、十二律（黄钟、林钟、太簇、南吕、姑洗、应钟、蕤宾、大吕、夷则、夹钟、无射、仲吕）与干支相匹配，推出六十甲子各自的五行属性，用来标明它们在不同阶段的性质特点。古人把六十甲子和纳音结合起来，就创立了《六十甲子纳音表》。

六十甲子纳音表，直观地诠释了《周易》的阴阳五行学说，对阴阳五行生克关系进行了推衍和拓展，把人类也按照金、木、水、火、土五类属性进行了划分，分成金命人、木命人、水命人、火命人、土命人。这样就为认识和把握人与人之间的相互作用关系，提供了强大有力的使用工具。

[1] ①朱伯崑：《周易知识通览》，齐鲁书社 1993 年版，第 167-177 页。

考虑到原表晦涩难懂，为了便于理解，本文把中国农历的时间和现今世界通用的公历进行相应换算，并标注了具体的公历年份，制作了《五类人命属性对照表》。见表 1。

表 1　五类人命属性对照表

出生年份	人命属性	出生年份	人命属性	出生年份	人命属性
1924/1984	金命人	1944/2004	水命人	1964/2024	火命人
1925/1985		1945/2005		1965/2025	
1926/1986	火命人	1946/2006	土命人	1966/2026	水命人
1927/1987		1947/2007		1967/2027	
1928/1988	木命人	1948/2008	火命人	1968/2028	土命人
1929/1989		1949/2009		1969/2029	
1930/1990	土命人	1950/2010	木命人	1970/2030	金命人
1931/1991		1951/2011		1971/2031	
1932/1992	金命人	1952/2012	水命人	1972/2032	木命人
1933/1993		1953/2013		1973/2033	
1934/1994	火命人	1954/2014	金命人	1974/2034	水命人
1935/1995		1955/2015		1975/2035	
1936/1996	水命人	1956/2016	火命人	1976/2036	土命人
1937/1997		1957/2017		1977/2037	
1938/1998	土命人	1958/2018	木命人	1978/2038	火命人
1939/1999		1959/2019		1979/2039	
1940/2000	金命人	1960/2020	土命人	1980/2040	木命人
1941/2001		1961/2021		1981/2041	
1942/2002	木命人	1962/2022	金命人	1982/2042	水命人
1943/2003		1963/2023		1983/2043	

表 1 说明：

1．每个人的五行属性，依据出生那一年的五行属性而定。中国的农历中，每一年都有各自的五行属性。因此，哪年出生的人，就与哪年的五行属性相同。例如，1924 年五行属性为金，这一年出生的人就是金命人；1926 年五行属性为火，这一年出生的人就是火命人。以此类推。由于每一年的五行属性各不相同，因此，每个人的五行属性也就不一样。

2．每一个周期相应年份的属性是相同的。中国农历每隔 60 年是一个循环周期，代表公元的纪年时间相差 60 年，因此，每一个周期相应年份的属性是相同的。也就是说，年龄相差 60 岁的两个人，五行属性是相同的。例如，1984 年和 2044 年，尽管之间相差 60 年，它们五行属性都为金，这一年出生的人都叫做金命人。以此类推。

3．表中计年的起讫时间是按照中国的农历来计算的。公历中起讫时间，是指每一年的 1 月 1 日起到 12 月 31 日止，而表中每一年的起讫时间，是从中国农历的节气“立春”那天开始，到下一年的“立春”之前结束。例如，1924 年立春前出生的人，只能认定为 1923 年出生的，这个人的五行属性不是金命而是水命。

三、研究思路、样本说明及计量方法

（一）研究思路

从上述《周易》理论中，本文在人力资本增值研究中得到了极大的启示：

1.任何事物或现象都存在规律性，人力资本的增值也不例外。各类人力资本在同一个企业总经理的领导下，增值的水平必然有高有低。而这种高低现象是可以通过人力资本的绩效考评结果显示出来的。假设 1，人力资本增值水平的这种高低现象，由于受到不同总经理的影响，会呈现出一定的规律性；假设 2，这种规律性可以通过阴阳五行生克关系原理的运用，在企业人力资本绩效考评的结果中找到，那么，依据这种规律性，对各类人力资本与总经理之间进行优化配置，就可以达到实现增值最大化的目的。

2．根据阴阳五行生克关系原理，以人力资本为主体，以总经理为客体，人力资本与总经理之间存在着 5 种关系：（1）相生关系。是指人力资本对总经理起着支持、帮助或促进作用；（2）相克关系。是指人力资本对总经理起着反对、排斥或制约作用；（3）相同关系。是指人力资本和总经理的人命属性相同，力量有所增强；（4）被相生关系。是指人力资本得到了总经理的认同、帮助或激励；（5）被相克关系。是指人力资本遭到了总经理的批评、压制或处罚。

因此，本文研究的思路是：（1）将各类人力资本以及总经理分别划分成五类人命属性；（2）分析研究各类人力资本在总经理影响下，5 年来 5 个方面的绩效考评结果呈现出的变化及其规律,以及对所在企业发展带来的不同影响；（3）对各类人力资本按照考评成绩的高低进行排序，分别建立最佳配置模型。

（二）样本说明

从 2008 年至 2012 年，本文研究采取问卷调查和深度访谈相结合的方法，在中国经济社会发展处于领先地位的江苏省南部 8 个县市（丹阳市、句容市、扬中市、金坛市、江阴市、宜兴市、常熟市、昆山市，都入选为中国基本竞争力 100 强县市），从中选择了 500 家企业，作为研究样本进行阐述。这些企业均属于亿元以上的民营企业，其中工业企业的占 94%以上，非工业企业的不到 6%。其中有 15 个国际知名品牌，86 个国家级品牌，有 91 个企业入选中国民营企业 500 强。为了分析研究的需要，本文以利税为主，对企业进行了划分：5 年来，平均年增长 20%以上的确定为超强势企业，10%—20%的确定为强势企业，10%以下的确定为弱势企业。根据划分，超强势企业有 139 家，强势企业有 263 家，弱势企业有 98 家。

本文研究的重点对象，是 500 家企业在册的人力资本。本文将人力资本主要分为三类：

1.核心人力资本，是指企业的副总经理、总监以及相当于这一职级的高层管理者。由于他们处于人力资本增值创造的核心地位，对企业的成功与失败具有决定性的作用，因此，对他们的配置就显得非常重要。

2.关键人力资本，是指企业各职能部门和生产车间（分厂、分公司）的主要负责人，是企业的中层管理者，也是一般人力资本的直接管理者，他们处于企业的关键岗位，对企业的人力资本和物质资本投资收益率都具有重大影响。考虑到职能部门和一线车间的职责有所不同，本文对关键人力资本分两种进行论述，并分别建立最佳配置模型。

3.一般人力资本，是指企业各职能部门的科员和车间包括班组长在内的一般员工。他们是企业经济保持持续健康发展的重要基础和保证。原本假想总经理对一般人力资本的绩效考评结果，会和核心和关键人力资本一样，产生很大的作用和影响，并有一定的规律性可寻，不料 2008 年、2009 年两次调查后，发现并非如此。2010 年第三次调查时发现，关键人力资本对一般人力资本的绩效考评结果，却有着决定性的作用和影响。因此，本文按

照一般人力资本工龄长短的顺序，采取等距抽样的问卷调查方法，对样本中不同类型企业的3000名在册员工，以关键人力资本为客体，作了一些统计学上的分析研究。见表2。

表2　各类人力资本人命属性及在企业分布情况

	人命属性						三类企业		
	金命人	木命人	水命人	火命人	土命人	小计	超强势	强势	弱势
核心	519	478	507	531	561	2596	701	1318	577
部门	874	1007	931	894	976	4682	1385	2454	843
车间	821	786	809	756	742	3914	1109	2040	765
一般	597	538	532	618	715	3000	1278	1022	700

据统计，500家企业的人力资本，职业年龄15年以上的达42%，10—15年的达36%，10年以下的达22%；核心和关键人力资本，大专以上学历分别达到了73%和41%，中级职称以上分别达到了54%和28%，而一般人力资本的大专以上学历和中级以上职称均不到20%，大多数是高中学历或初级职称，表明一般人力资本对自身投资以及所在企业对其专业培训缺乏足够的重视。

（三）计量方法

本文研究中采用的数据，来源于500家企业每年对人力资本的绩效考评结果。目前中国企业通用的绩效考评内容，是参照欧美发达国家企业绩效考核管理的模式，在工作分析的基础上，对人力资本在道德品质（工作态度）、学识水平、工作能力、工作业绩、性格素质等5个方面进行评价，以此作为奖励、处罚、培训、辞退、职务任用等实施的基础与依据。实践证明，这是一种比较科学的程序和方法。

考评一般采取定性和定量相结合的方法。通常，企业核心和关键人力资本的考评结果，按照高低排序分为4个不同档次，档次越高表示贡献越大，档次越低表示价值越小。对一般人力资本进行考评，是将上述5个方面因素用分值或系数的形式来表示，一般也分成四个档次：90分以上、80分—90分、70分—80分、70分以下。本文统一换算成定性方式，用“优秀、良好、合格、不合格”来作计量分析。

四、实证检验与结果分析

（一）实证检验

按照研究思路，为了检验假设1、假设2是否成立，本文将各类人力资本，在5年时间里（一般人力资本只有3年），在5个考核内容规定的方面，所获得 “优秀”、“良好”、“合格”和“不合格”的人数及比率，进行平均计算后得出了总体考评结果，然后，按照阴阳五行生克关系，将各类人力资本与领导自己的总经理一一对应、归类（一般

人力资本与关键人力资本相对应）。因篇幅有限，本文未能列出关键人力资本、一般人力资本的绩效考评总体成绩，仅能提供核心人力资本绩效考评总体成绩用于分析描述。见表3。

表3　核心人力资本绩效考评5年平均人数及比重

总经理 / 核心资本		总经理人命属性									
		金命人		木命人		水命人		火命人		土命人	
		人数	比重	人数	比重	人数	比重	人数	比重	人数	比重
金命人	优秀	44.80	42.67	16.56	22.68	58.28	49.39	37.60	27.45	31.80	36.98
	良好	42.04	40.03	21.44	29.37	42.72	36.20	42.92	31.33	31.60	36.74
	合格	14.16	13.49	20.04	27.45	14.40	12.20	34.12	24.91	14.76	17.16
	不合格	4.00	3.81	14.96	20.49	2.60	2.20	22.36	16.32	7.84	9．12
木命人	优秀	31.76	29.14	49.60	51.67	42.64	35.53	58.64	58.06	13.60	26.15
	良好	36.52	33.50	29.80	31.04	44.48	37.07	29.00	28.71	13.08	25.15
	合格	24.72	22.68	12.92	13.46	22.64	18.87	12.24	12.12	12.64	24.31
	不合格	16.00	14.68	3.68	3.83	10.24	8.53	1.12	1.11	12.68	24.38
水命人	优秀	34.76	38.62	50.24	60.53	52.16	53.77	30.24	23.08	30.44	28.72
	良好	32.88	36.53	22.76	27.42	29.24	30.14	35.12	26.81	36.88	34.79
	合格	17.40	19.33	8.64	10.41	12.64	13.03	32.12	24.52	23.16	21.85
	不合格	4.96	5.51	1.36	1.64	2.96	3.05	33.52	25.59	15.52	14.64
火命人	优秀	18.96	24.00	44.24	35.39	34.48	33.15	39.40	45.81	65.92	48.12
	良好	21.24	26.89	45.40	36.32	31.28	30.08	32.08	37.30	53.24	38.86
	合格	20.32	25.72	23.32	18.66	24.80	23.85	12.16	14.14	15.64	11.42

	不合格	18.48	23.39	12.04	9.63	13.44	12.92	2.36	2.74	2.20	1.61
土命人	优秀	61.12	47.38	33.08	26.89	10.80	18.00	42.68	34.98	49.08	38.65
	良好	51.28	39.75	39.28	31.93	14.80	24.67	46.16	37.84	53.72	42.30
	合格	14.80	11.47	32.16	26.15	17.12	28.53	25.68	21.05	20.64	16.25
	不合格	1.80	1.40	18.48	15.02	17.28	28.80	7.48	6.13	3.56	2.80

研究发现，各类人力资本所获得的总体考评结果，在五行生克关系的作用下，不是杂乱、无序的，而是呈现出一定的规律性。这种规律性具体表现为：

1.生克关系的客体对主体人力资本的考评优劣具有决定性影响

核心和关键人力资本作用的客体是总经理，一般人力资本作用的客体是关键人力资本，他们之间彼此存在不同的五行生克关系。在客体的领导下，受阴阳五行生克关系的影响，各类人力资本在5个方面的绩效考评成绩或优或劣，每年在平均分值与人数比重的高低上，都会呈现出规律性的排序现象。高的总是高，低的总是低。

表3显示，以核心人力资本获得“优秀”的比重为例：

（1）当是金命人时，无论哪一年的考评，在水命人总经理手里得到的“优秀”比重总是最高，在木命人总经理手里得到的总是最低。

（2）当是木命人时，无论哪一年的考评，在火命人总经理手里得到的“优秀”比重总是最高，在土命人总经理手里得到的总是最低。

（3）当是水命人时，无论哪一年的考评，在木命人总经理手里得到的“优秀”比重总是最高，在火命人总经理手里得到的总是最低。

（4）当是火命人时，无论哪一年的考评，在土命人总经理手里得到的“优秀”比重总是最高，在金命人总经理手里得到的总是最低。

（5）当是土命人时，无论哪一年的考评，在金命人总经理手里得到的“优秀”比重总是最高，在水命人总经理手里得到的总是最低。

部门、车间关键人力资本和一般人力资本也是如此。随着与客体之间所处的五行生克关系不同，而呈现出相当大的差距：凡是相生的，得到“优秀”、“良好”的机会和概率就多；反之，凡是相克的，“合格”、“不合格”的机会和概率就会明显增加。

根据各类人力资本考评成绩的高低，可以相应得出增值大小的排序位次：

（1）核心人力资本是：相生>相同>被相生>被相克>相克；

（2）部门关键人力资本是：相生>被相生>被相克>相同>相克；

（3）车间关键人力资本是：相生>被相克>相同>被相生>相克。

（4）一般人力资本是：相生>相同>被相生>被相克>相克。这与核心人力资本的排序完全一样，只是作用的客体不同而已。

2.人力资本优劣对企业效益差异具有决定性影响

对企业而言，每年能够获得优秀比重最多的人力资本，无疑是能够实现增值最大化的人力资本。一个人力资本自身获得的优秀比率多少，究竟会给所在企业带来怎样的影响？

本文将各类人力资本的总体考评结果，与他们所在的企业相互对应，分析发现人力资本的优劣对所在企业的效益差异具有决定性影响。限于篇幅，本文未能提供部门、车间的关键人力资本和一般人力资本的总体考评成绩在三类企业的分布情况，仅以核心人力资本总体考评成绩在三类企业的分布情况进行描述。见表 4。

表 4　核心人力资本总体考评结果在三类企业的分布

核心资本 / 企业类型		核心人力资本							
		优秀	比重%	良好	比重%	合格	比重%	不合格	比重%
超强势	金命人	71.60	48.71	52.28	35.56	18.20	12.38	4.92	3.35
	木命人	58.28	50.68	44.48	38.68	8.12	7.06	4.12	3.58
	水命人	60.24	48.58	45.16	36.42	15.24	12.29	3.36	2.71
	火命人	77.68	51.79	51.52	34.35	15.28	10.19	5.52	3.81
	土命人	83.36	50.52	55.32	33.47	19.56	11.85	6.76	4.10
强势	金命人	100.32	34.24	106.4	36.31	56.16	19.17	30.12	10.28
	木命人	103.84	41.21	86.64	34.38	38.28	15.19	23.24	9.22
	水命人	99.96	39.51	89.32	35.30	39.08	15.47	24.64	9.74
	火命人	93.24	38.06	96.08	39.22	41.56	16.96	14.12	5.76
	土命人	83.28	30.28	120.16	43.69	51.12	18.59	20.44	7.43
弱势	金命人	17.12	21.67	22.04	27.90	23.12	29.27	16.72	21.16
	木命人	34.12	30.74	21.76	19.60	34.76	31.32	14.36	12.94
	水命人	37.64	28.95	22.40	17.23	39.64	30.49	30.32	23.32
	火命人	32.08	23.59	35.64	26.21	39.40	28.97	28.88	21.24
	土命人	30.12	24.89	29.76	24.60	39.72	32.83	21.40	17.69

研究发现，在五行生克关系的作用下：

（1）超强势企业中，各类人力资本的优秀人数占有比率，明显高于其它两类企业，而不合格的人数占有比率，明显低于其它两类企业；

（2）弱势企业中，各类人力资本的优秀人数占有比率，明显低于其它两类企业，而不合格的人数占有比率，明显高于其它两类企业。按照超强势、强势和弱势进行排序，各类人力资本的优秀率依次呈下降状态，而不合格率依次呈上升状态。

以火命人的核心人力资本为例，其优秀人数占有比率：超强势企业高达 51.79，强势企业中占到 38.06，而弱势企业中只有 23.59，比强势企业少了 14 个百分点，还不到超强势企业的 1/2；其不合格占有比率：超强势企业只有 3.81，强势企业有 5.76，而在弱势企业中却高达 21.24，分别是超强势、强势企业的 5 倍和 3 倍以上。

（二）结果分析

上述规律性产生的原因，对现代科学来说一时难以解释，但因此使配置的价值意义，在阴阳五行学说的承托下，从企业人力资本运营的深处浮出了水面。研究表明，配置的好坏，直接关系到人力资本的增值大小和企业运营的成功失败。

1. 配置对人力资本增值具有决定性的作用。每个人力资本都具有与生俱来的价值意义，但是在同等努力的条件下，由于受到参照对象在五行生克关系上的影响，实现的增值大小就不同。配置合适了，增值就大，反之，增值就小。可见实现人力资本增值最大化关键在于优化配置。通过优化配置，使人力资本与参照对象之间的人命属性关系更合适、更融洽、更和谐，是实现人力资本增值的最大化的有效路径。

2. 企业经济增长主要来源于人力资本的价值增值。从管理角度上讲，人力资本之间不存在优劣之别，但在某个时空点上会有增值上的差别。研究表明，企业配置优化了，人力资本的优秀比率就高，实现增值的程度就大，就会发展成为超强势企业；反之，如果不能发挥好或无视配置的积极作用，人力资本难以保值增值，企业就会出现发展停滞或利润下滑的痛苦局面，最后衰败成了摇摇欲坠的弱势企业。

3.企业配置的实质，就是对人与人之间外在环境相互作用关系的调整和处置。如何做到优化配置，五行生克关系原理已经为我们提供了方法论。是否科学，有待于今后进一步研究。但是，在当今人力资本配置研究中，还找不到比其更有效的方法论。因此，应用五行生克关系的作用机制，是对人力资本进行优化配置最直接、最有效的手段，必将成为企业的唯一选择。企业只要做到优化配置，人力资本运营实现收益最大化，就会从理想变为现实。

五、配置模型的建立

根据上述各类人力资本绩效考评成绩由高往低的排序，由此建立最佳配置模型。

5　相生
4　相同
3　被相生
2　被相克
1　相克

5　相生
4　被相生
3　被相克
2　相同
1　相克

图 1　核心人力资本配置模型

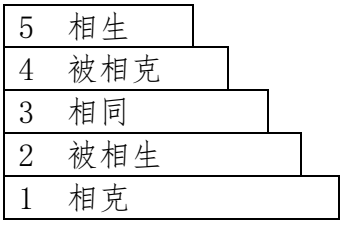


图 2　部门关键人力资本配置模型

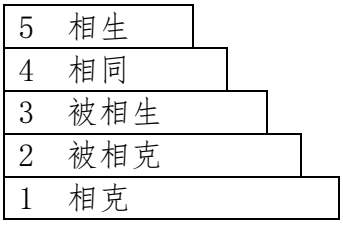


图 3　车间关键人力资本配置模型

图 4　一般人力资本的配置模型

模型应用说明：

1.图中用“5、4、3、2、1”分值，来表示人力资本价值的高低，分值越高，资本价值越高，反之，分值越低，资本价值也越低。分值必须大于或等于 3，各类人力资本才有保值增值的可能性。

2．当对人力资本个体进行增值评价或计划招聘时，具体操作步骤是：（1）先到表 1 中分别找出总经理（客体）和各类人力资本的人命属性；（2）按照五行生克关系进行对照，以各类人力资本为主体，确认总经理（客体）与各类人力资本之间的相互关系；（3）对照配置模型，就知道哪个人力资本能否实现增值的最大化，或是企业选聘的人力资本是否优秀。例如，某个企业的总经理，是 1964、1965 年出生的火命人，那么，核心人力资本中，第一人选是相生总经理的木命人，第二人选是与总经理相同的火命人，第三人选是土命人，不合适的人选是金命人和水命人。部门、车间和一般关键人力资本的增值评价或计划招聘，具体操作步骤也参照上述方法进行。值得注意的是，一般人力资本的客体是关键人力资本，应用时切忌混淆。

3.根据上述配置模型，画成一张各类优秀人力资本招聘表。利用这张表，每个企业就可以有目标、有计划地去选聘各类优秀的人力资本，同时，每个亟待就业的人力资本也可以找到适合自我发展、展示人生价值的企业。见表 5。

表 5　企业优秀人力资本招聘表

总经理 / 资本类型		总经理（客体）				
		金命人	木命人	水命人	火命人	土命人
核心资	金命人	4	1	5	2	3
	木命人	2	4	3	5	1

本	水命人	3	5	4	1	2
	火命人	1	3	2	4	5
	土命人	5	2	1	3	4
部门资本	金命人	2	1	5	3	4
	木命人	3	2	4	5	1
	水命人	4	5	2	1	3
	火命人	1	4	3	2	5
	土命人	5	3	1	4	2
车间资本	金命人	3	1	5	4	2
	木命人	4	3	2	5	1
	水命人	2	5	3	1	4
	火命人	1	2	4	3	5
	土命人	5	4	1	2	3
一般资本	金命人	4	1	5	2	3
	木命人	2	4	3	5	1
	水命人	3	5	4	1	2
	火命人	1	3	2	4	5
	土命人	5	2	1	3	4

六、结 语

本研究对人力资本理论的贡献，还在于创新了企业人力资本增值研究的视角。对“企业如何实现人力资本增值的最大化”这一世界性难题作了非常有益的探索，为各类人力资

本在择业上做出更加有利于自身发展的抉择指明了前进方向，为企业优化配置提供了一个实用性强的使用模型，开辟了人力资本实证研究的崭新思路。研究中还发现，配置模型可以移植到国有企业或社会（政府）组织结构，一旦全面实施将会产生革命性的意义。

值得注意的是，本文研究还存在着一定的局限性：由于条件限制，样本数据在时间跨度上只有 5 年，在经济区域上主要立足于经济发达的中国苏南地区，某种程度上可能会影响统计数据检验的效力；配置模型对各类人力资本的增值进行了定性，而在数值的测度上还存在着明显不足，未能作出进一步的解释，需要继续进行大量的实证研究。

科学研究不仅是无国界的，更是无时空的。古代文明的智慧是老祖宗留给我们子孙后代的一笔宝贵的财富。在现代科学技术发展的今天，它依然闪烁着耀眼的光芒，给后人以深刻的启示，我们只要发掘、认识和利用好，就能使人类社会拥有更大的合作和发展空间。

作者简介：　刘绛昊（1965—？），男，江苏丹阳人，曾任江苏省丹阳市行政学校常务副校长 12 年，现在丹阳市人民政府法制办公室工作，长期从事经济学、哲学和《周易》经学的研究。联系地址：江苏省丹阳市开发区兰陵路 8 号　邮政编码：212310　座机：0511—86922555
电子信箱：576864026@qq.com　　手机：13806108398。

关于感染组学研究

曹虹

南方医科大学公共卫生学院微生物学教研室

二十一世纪人类仍将面临各类传染病（包括新现和再现传染病）以及生物恐怖主义的严峻挑战。细菌、病毒、真菌和寄生虫性病原体及其产物所引起的感染性疾病迄今仍然是全球范围内高发病率和高死亡率的疾病之一。由于上述病原体的感染而引发的当今医学领域另一严重问题是，广泛、大量和不恰当地使用抗生素导致耐药病原体的传播和扩散，其最终后果将使人类返回到抗菌素发现前的时代。另外，在临床治疗中大量采用干预疗法及延长病人寿命，使老龄患者和免疫受损的病人不断增多，大大增加了致命性条件感染的机率。在全球恐怖活动日益猖獗的今天，生物恐怖袭击的威胁已越来越现实。如何防止生物恐怖也是摆在我们面前的严峻问题。因此，发展新的抗感染手段对付上述新现和再现、耐药和突变的病原体已成为现代医学的当务之急。

自从人类基因组计划完成人类基因组解码后，诸多新兴学科如基因组学、蛋白组学和糖组学等“组学（omics）”不断涌现，其共同特点是采用“整体观”（holism）揭示生命现象的本质。正是由于这些新兴学科在方法论上的革新，才为人类应对上述挑战提供了机会。感染组学（Infectomics）是一门新学科，它是对病原性微生物感染的“组学”研究。包括两个主要方面：结构研究和功能研究二个方面。自 2002 年黄胜和教授首次提出感染组学（Infectomics）的概念以来，这一领域的研究在国际上备受关注，已被 Nature《人类基因组百科全书》列为三大医学组学之一。

感染组(infectome)是指当机体受到感染时所引起的基因型和表型的“组学”改变；而这些改变是由病原微生物和宿主的基因组编码，并且包括了病原体及感染的宿主体内的基因复制（DNA）、转录(mRNA)、翻译（蛋白质）和翻译后修饰（如糖基化和磷酸化）等水平上的变化。在现代医学界，感染组的特征反映病原体和宿主之间的相互作用，为研究微生物致病机理，探索预防病原体感染的新方略及发展新的抗微生物药物提供了极其丰富又有价值的资料。

目前，在感染性疾病的范畴，急需一个更有效、更精确和全面综合性地研究病原微生物感染的结构和功能基因组学和蛋白质组学，即感染组学的方法。新方法（如 DNA 和蛋白质微阵列）和传统方法（如分子克隆、PCR、基因敲除和加进等）的结合将有助于克服今天的困难。感染组学不但能大大革新研究微生物致病机制的方法，而且可以创新对感染性疾病的处理、对未知病原体的认识、解决目前对抗生素抗性的危机、扩大疫苗候选物和研制具有双重靶向性的抗感染药物。对病原微生物引起的感染性疾病的预防和治疗亦将最后进入整体性预测与个性化处理的时代。

美国南加州大学黄胜和教授是目前国际上感染组学研究的权威，为自然杂志社(Nature)出版的《人类基因组百科全书》一书“感染组学”部分的撰稿人，同时亦为 Journal of Molecular Biology and Biotechnology (JMBAB) 编委，和 Cell Microbiology、FEBS Journal (Eur J Biochem)、Drug Discovery Today、Clin Infect Dis、J Infect Dis、Microb Pathogenesis、 Mol Microbiology 以及 Nucleic Acid Research 杂志的阅稿人，与南方医科大学微生物学教研室和华中理工大学有长期的合作研究关系。2010 年黄

教授提出的用非线性焦点博弈论模型研究 HIV/AIDS 病已获得美国盖茨研究基金奖。最近黄教授为《抗感染药物》杂志主编了感染组学(Infectomics) 专题特刊 。在该特刊中，首次提出以内共生(Endosymbiosis) 与外共生(Exosymbiosis) 平衡(EESB)为基础的生态免疫感染组学理论。

References

[1]. S.H. Huang, T. Triche and A.Y. Jong, “Infectomics: genomics and proteomics of microbial infections,” Functional & Integrative Genomics. Vol.1, pp.331-44, 2002.

[2]. Huang SH, Jong, AY, Summersgill, JT. Infectomic Analysis of Microbial Infections Using Proteomics, In Handbook of Proteomics, Humana Press, 2003, pp.347-354.

[3]. Huang SH. Infectomics: the study of response to infection using microarrays, In Encyclopedia of the Human Genome (Editor-in-Chief: D.N. Cooper), Nature Publishing Group, 3:453-458, 2003.

[4].Huang, SH, Jong, AY. Infectomics Overview: Holistical and Integrative Studies of Infectious Diseases. In Techniques in Molecular Analysis and Genome Discovery (Edited by R. Rapley). John Wiley & Sons, Ltd, P.253-272, 2004.

[2]. 黄胜和 徐钤. 感染组学(infectomics) - 生物化学与生物物理进展. 32:304-9, 2005.

[3]. 黄胜和 徐钤. 感染组学在探索新的抗微生物药物中的应用 - 科技导报. 23:44-48, 2005.

[4]. S.H. Huang, X. Wang and A. Jong, “The evolving role of infectomics in drug discovery,” Expert Opinion in Drug Discovery, Vol. 2, pp.1-15, 2007.

[7]. Huang SH, Li BL, Cao H. Infectomics: study of microbial infections using omic approaches. Encyclopedia of Life Sciences. 2007, 1-8.

[5]. Huang SH. Drug discovery and development using infectomics and related approaches: Introductory Overview for Anti-Infective Agents (Editorial), Anti-Infective Agents, 11:91-92, 2013.

[6]. Huang SH, Tang SQ, Bo T, Xia HW, Jong A. Ecological infectomics (Ecoinfectomics): symbiotic solutions to microbial infections, Anti-Infective Agents, 11:93-99, 2013.

[7]. Zhu QQ, Wang L, Tang SQ, Cao H, Huang SH. Exploring novel immunological interventions using immunoinfectomics, Anti-Infective Agents, 11: 109-114, 2013.

[8]. Huang SH. Xia HW, Chi F, Khalaf NK, Jong A. Chemoinfectomics in drug design and development, Anti-Infective Agents, 11: 100-108, 2013.

[9]. Yang ZW, Jiang YH, Ma C, Silvestri G, Bosinger SE, Li BL, Jong A, Zhou YH, Huang SH. Coexpression Network Analysis of Benign and Malignant Phenotypes of SIV-infected Sooty Mangabey and Rhesus Macaque. PLoS One. 11:e0156170.2016.

[10]. Huang SH, Zhou YH. Infectomic Insights into the Roles of Exosymbiosis-Endosymbiosis Imbalance in HIV-1 Infection. J Data Mining Genomics Proteomics. 7:196, 2016.

竞争中立问题研究

丁茂中
上海政法学院竞争法研究中心

内容摘要：竞争中立如今已经嬗变成为一个涉及国内与国际、法律与政治、理念与手段的复杂问题。罗伯特·霍马茨之谬与澳式竞争中立的误传表明，作为普世理应信奉与践行的竞争法治理念，竞争中立的科学含义应为：在市场有效竞争的环境下，除基于克服潜在的市场失灵需要而依法进行非中立性的干预之外，政府应当公平地对待参与市场竞争的所有企业。在竞争中立的本义下，一国政府应当在交易机会、经营负担、投资回报三个竞争点面上保持中立，但是非市场化领域的政府管理、引入竞争机制的政府改革、维持竞争机制的政府规制不适用竞争中立的基本行为准则。各国应当通过综合采取竞争法律规制、国贸协定约束与竞争文化引导治理政府在干预市场过程中出现的滥用权力限制、排除竞争问题。

关键词：竞争中立　市场法治理念　贸易保护手段　制度规则　治理路径

引言

竞争中立，按照媒体所言，这一观念是美国国务院负责经济、能源和农业事务的副国务卿罗伯特·霍马茨（Robert Hormats）不久前在纽约外国记者中心举行的发布会上提及的概念，意思是使竞争不受外来因素的干扰，其核心是对现有国际经济规则进行更新和调整，以“弥补现有的国际经济规则无法保证国有企业和私营企业公平竞争的缺陷”；[1]但是罗伯特·霍马茨本人却强调：竞争中立一词并非其杜撰出来的，这一包罗广泛的主题已经涵盖在《泛太平洋战略经济伙伴关系协议》（TPP）这一贸易协议当中，而且在经济合作与发展组织（OECD）的工作当中有更具体的体现。[2]然而，不管竞争中立的“身世”真相究竟如何，但是一个现实结果就是经过罗伯特·霍马茨在各种重要场合下三番五次的提及使得竞争中立逐步与政策、框架等联系在一起成为中国官方与民间社会各界关注的重大问题（以下统称竞争中立问题）。社会公众对竞争中立问题的认识较为直观，很多人应情应景地引入竞争中立对我国目前仍然客观存在的一些政府差别对待国企与民企现象针砭时弊；而政府高层和部分专家学者对竞争中立问题的认识则较为深远，他们在充分意识到社会公众所聚焦的“内忧”同时更多的开始着眼思考中国经济发展将要面临的一个重大“外患”，即美国蓄谋利用比现行世界贸易组织（WTO）框架下任何规则杀伤力都要强数倍的竞争中立问题实施贸易保护。[3]比较“内忧”与“外患”，毫无疑问，道理上后者应当是现阶段中国经济发展的主要矛盾之一。但是因竞争中立的字面意思极度富有哲理性、我国长期以来客观存在的国企民企有别现象所导致的社会反感情绪、美国主导的TPP所推出的竞争中立框架和美国“领导”的OECD所倡导的竞争中立政策共同夹带的不纯动机具有较强的隐蔽性，我国很多人目前并没有真正及时明白这个道理，他们过度放大“内忧”或者浑然不知“外患”而将前者视为现阶段中国经济发展的主要矛盾之一。这正中美

[1] 吴云：《美国出新招对付“中国模式”把中国国有企业竞争力当靶子》，《人民日报》2011年11月24日。
[2] [美]博伯·戴维斯：《美国副国务卿：中国国企不应当扭曲竞争》，《中国实时报》2011年11月28日。
[3] 参见《为实施贸易保护“竞争中立”登台亮相》，《中国产经新闻》2012年2月6日。

国力推的新全球经济游戏规则下怀，将会给我国经济乃至政治、军事、文化等的发展产生巨大的负面影响。因此，全面深度解析竞争中立问题是非常必要的；这不仅有助于我们科学地采取措施抵御“外患”，而且有助于我们科学地采取措施消除“内忧”。

一、竞争中立问题的性质嬗变

若要明察秋毫现代各种语境之下的竞争中立问题，我们首先必须站在市场经济历史的高点弄清竞争中立问题的全貌。因适用范围、博弈群体、援用目的等悄然发生巨大变化，竞争中立早已远远超出人们常规理解的字面意思而演变成为一个错综复杂的问题。

（一）国内问题到国际问题

市场固有的地域属性决定了经济领域的系统竞争最早成形于各个国家，这使得竞争中立问题初始只是一个国家的国内问题。从实践来看，虽然很多国家先前在讨论、制定、执行本国的相关经济政策、法律规范等过程中客观上实质性地考虑了竞争中立问题，但是率先明确提出竞争中立问题的却是澳大利亚。早在 1996 年，澳大利亚就发布了《联邦竞争中立政策声明》（Commonwealth Competitive Neutrality Policy Statement），首次提出并定义了竞争中立概念，即政府商业活动不应当仅凭其公共部门所有权而享有高于私营部门竞争者的竞争优势。[4] 2004 年，澳大利亚国库部和财政部联合发布了《澳大利亚政府对经理人的竞争中立指引》（Australian Government Competitive Neutrality Guidelines for Managers）；在该指引文件中，澳大利亚特别声明：政府致力于令人信服、透明的行政管理和公共资源的高效配置与使用，实现此承诺的一个重要路径就是对政府商业活动实行竞争中立政策，范围涵盖税收、信贷、政府管制、投资回报、成本分担、诸如保险等其他领域等。[5]

但是随着市场经济走向全球化，竞争中立问题逐步突破主权篱笆成为国际关注的一个重要问题。从形式上来看，国际对竞争中立问题的关注来自三个层次：第一，影响全球的大国，如美国。除政府高官在各种场合提及竞争中立问题外，美国目前实实在在地采取了很多措施竭力推进 TPP 的谈判进程。“TPP 不仅要求减低或取消关税，更重要的是它强调各成员国的竞争政策，例如推行高标准的知识产权保护、金融改革、增强规则透明度等一系列措施；实际上就是强调规范政府与市场的关系，推行竞争中立政策。”[6] 第二，区域性国际组织，如欧盟。[7] 除了与美国共同发表《关于国际投资共同原则的声明》（Statement of the European Union and the United States on Shared Principles for International Investment），强调支持 OECD 在竞争中立领域所做的工作外；[8] 欧盟还在其区域性法律规范中对竞争中立问题作了一些实质性规定，如《欧共体条约》第 86 条第 1 款规定，成员国不得对国有企业以及享有特权或专有权的企业采取背离或者保留本条约，特别是条约第 12 条以及第 81 条至第 89 条的任何措施。第三，全球性国际组织，如 OECD。OECD 是最早研究竞争中立问题的全球性国际组织，在 2001 年—2010 年先

[4] Australian, Commonwealth Competitive Neutrality Policy Statement, http://archive.treasury.gov.au, 2013.08.17.

[5] Australian, Australian Government Competitive Neutrality Guidelines for Managers, http://www.finance.gov.au, 2013.08.17

[6] 赵学清、温寒：《欧美竞争中立政策对我国国有企业影响研究》，《河北法学》2013 年第 1 期。

[7] 因欧盟在经济、法律、政治等方面高度一体化，本文沿袭惯例未严格区分欧盟在主体性质上的使用问题，故在有些“国家”语境下亦会提及欧盟相关方面内容。

[8] EU&USA, Statement of the European Union and the United States on Shared Principles for International Investment, http://trade.ec.europa.eu, 2013.08.20.

后发布一系列诸如《OECD 关于财政预算透明度最佳实务》（OECD Best Practices for Budget Transparency）、《OECD 关于国有企业公司治理的指引》（OECD Guidelines on Corporate Governance of State-owned Enterprises）等文件后，OECD 于 2012 年发布了《竞争中立：维持公有企业与私有企业之间竞争水平》（Competitive Neutrality: Maintaining a level playing field between public and private business），该文件重点对简化国营企业运营方式、特定功能的成本核算、获得商业回报率、提供公共服务、税收中立、政府管制中立、信贷中立及合计补贴、公共采购等八个方面的竞争中立问题作了阐述。[9]

（二）法律问题到政治问题

一国对国内市场经济活动中竞争中立问题的重视通常都是基于本国竞争法律制度的有效实施，虽然竞争中立问题的彻底解决往往涉及到政治、经济、文化等诸多方面的改革，但是市场经济的法治化使得国内竞争中立问题的处理在整体上还是置于法律范围内。作为竞争中立的先驱者，澳大利亚在竞争中立问题的起源报告《国家竞争政策》（即著名的希尔墨报告）中阐述道：“每一个现代市场经济都有一套规则确保竞争机制不会被企业通过合谋或者单干的反竞争行为破坏。…在澳大利亚，这些规则包含在《1974 联邦商业行为法》（Commonwealth Trade Practices Act 1974）第四部分。…当前最紧迫的问题就是确保这些规则在适用过程中所存在的不合理漏洞被弥补上以给企业经营提供一个全国统一的法律框架。”[10]为了解决体制性问题给市场竞争带来的诸多不公平问题，澳大利亚联邦与新南威尔士州、昆士兰州等八个地方主体签订了《竞争原则协议》（Competition Principles Agreement），并在《联邦竞争中立政策声明》中建立了一套以生产力委员会为主要执法机构的纠纷解决法律机制。

但是国际对跨国市场经济活动中竞争中立问题的关注比较复杂，考量内容很多都已超出纯粹的法律问题范畴而带有浓烈的政治色彩。首先，美国、OECD、欧盟在竞争中立问题上的态度所呈现的遥相呼应表象本身就是美国精心策划的政治布局。美国主导的 TPP 谈判进程因竞争中立问题进展十分迟缓，这使得美国在竞争中立问题上迫切需要他方强有力的舆论支持，而 OECD 和欧盟则是理想人选。OECD 成员的地域广泛性和标签的经济主导性使得 OECD 发布的文件通常都会有着较高的国际影响力，因作为创始国而具有主导掌控力的美国自然不会浪费利用 OECD 的机会，这也就是 OECD 近年来不断发布《OECD 关于财政预算透明度最佳实务》、《OECD 关于国有企业公司治理的指引》、《竞争中立：维持公有企业与私有企业之间竞争水平》等多个富有逻辑连贯性文件的重要幕后原因。作为全天候全方位的盟友，虽然欧盟只是与美国共同发表了《关于国际投资共同原则的声明》，但是因欧盟市场的全球重要性和竞争规则的法治化，欧盟在竞争中立问题上的表态对于一定程度打消国际多方对于美国推行竞争中立的种种疑虑是具有非同寻常的声援意义。其次，美国以“弥补现有的国际经济规则无法保证国有企业和私营企业公平竞争的缺陷”为藉口竭力推行竞争中立隐藏着深刻的政治用意。自进入二十一世纪以来，伴随着中国、巴西、俄罗斯、印度等国家的飞速发展，以国家主导的产业政策为核心的国家资本主义模式逐渐引起很多国家的关注并逐步成为联合国贸易和发展会议（UNCTAD）的重要议题之一。国家资本主义模式的迅速兴起不仅开始威胁到自由资本主义模式在全球经济领域的霸主地位，而

[9] OECD, Competitive Neutrality: Maintaining a level playing field between public and private business, http://www.oecd.org, 2013.08.20

[10] Australian, National Competition Policy, http://ncp.ncc.gov.au, 2013.08.24.

且直接影响到美国在全球经济乃至政治领域的领导地位。为了扭转不利境况、巩固领导地位，美国以 TPP、OECD、欧盟为竞争中立问题的舆论支撑基点通过四处宣扬所谓的中国模式“扭曲”了竞争（罗伯特·霍马茨用语）的方式图谋达到凸显国家资本主义模式缺乏基本的法理基础、全面遏制中国日益扩大的国际影响力的政治目的，并藉此将竞争中立问题突破 TPP 谈判范畴推向全球化以谋求建立一套主要有利于美国的新国际贸易规则。

（三）理念问题到手段问题

思想发源于现代市场经济的竞争中立初本仅为一国竞争法治的一种理念，即政府应当公正对待主权辖区内所有纳税的经营者以给他们提供一个公平竞争的机会，特别是在国有企业与私营企业之间。纵观澳大利亚竞争中立政策发展史，以竞争中立作为市场竞争法治理念的实践脉络是非常清晰可见的。1974 年，澳大利亚通过《1974 联邦商业行为法》取代了过时的《1965 联邦商业行为法》。但是根据当时的条文来看，这部新法仍然存在很多问题，如无法适用于合伙、个体商贩等非法人企业之间在州内的商业行为、联邦和州政府制定的其他法律所规定的适用除外条款仍然具有优先效力等。这不仅导致澳大利亚的政府商业活动与私营部门之间依然存在诸多不公平竞争问题，而且导致澳大利亚不同领域的私营部门竞争者之间出现不公平竞争问题。为了彻底改善澳大利亚的竞争法治环境，联邦和各州在 1991 年决定成立一个独立的国家竞争政策检查委员会全面调查澳大利亚的市场竞争状况，该委员会在 1993 年公布了前面提到的著名希尔墨报告（The Hilmer Report）。这份报告引起了澳大利亚政府委员会的高度重视，随即在 1995 年澳大利亚联邦与各州一致同意通过了《竞争原则协议》对《1974 联邦商业行为法》所存在的上述一系列问题作了修正。为了进一步充分彰显澳大利亚的竞争中立市场竞争法治理念，澳大利亚除后续不断采取各种措施深化竞争中立改革之外，[11] 它还在 2011 年将《1974 联邦商业行为法》更名为《2010 竞争与消费者法》（Competition and Consumer Act 2010）。

但是当今国际层面倡导的竞争中立在很大程度上已经变味沦为大国政治博弈的一种手段，即以形式上同等适用于自身的竞争中立规则限制他国政府采取特定措施对国际市场竞争中现存的在实质上属于不公平竞争的问题进行矫正以借此维持本国企业在国际市场中的优势地位。揭开美国到处推行竞争中立的面纱，美国惯性操弄游戏规则损人利己的本性一览无遗。因两次世界大战等特殊因素而造就的世界领导地位使得美国企业因货币优势、技术优势、资本优势、规则主导优势等在国际市场竞争中一直处于领先地位，这对于逐步走向与之同台竞技的广大发展中国家的企业而言在实质上是非常不公平的。为了改变在世界经济结构中的被动劣势地位，包括中国在内的很多发展中国家纷纷利用产业政策集中资源重点发展高新技术产业、相对优势产业等。中国经济的突飞猛进使得美国对于这种政府干预市场模式非常忧虑甚至恐惧，但是因现行 WTO 框架对此是“鞭长莫及”的，所以美国竭力尝试“变法”伪善地推出了竞争中立，企图借此捆住广大发展中国家特别是中国的政府产业政策之手以维持美国及其企业的现有地位。有些行业专家更是进一步指出：“美国推动竞争中立，主要是以中国等国家普遍存在的国有企业为目标，在美国与其他国家的竞争中保护美国的利益。一方面，美国积极推动他国的自由贸易，从而为本国商品和资金打开

[11] 澳大利亚后续采取的措施参阅生产力委员会等发布的文件，详细内容可登陆澳大利亚的国家竞争政策官网（http://ncp.ncc.gov.au/pages/home）“主要改革领域”（Major Areas of Reform）和“出版检索”（Publication Search）两个栏目查看。

市场；另一方面，美国又加强对本国产业的保护，抑制别国竞争。此外，美国也想借此继续抑制其他国家通过投资获取美国技术。”[12]

竞争中立问题的错综复杂意味着我们在引用“他山之石”进行“攻玉”的过程中不能仅凭感性的文义认知进行作为，必须理性地透过表象把握竞争中立的本质根据自身的整体发展需求科学地作出择用。

二、竞争中立的科学含义

性质的嬗变使得竞争中立问题在国际层面与国内层面明显发生了“位移”，但是一些国际主体特别是美国所倡导的竞争中立貌似又“持之有故、言之有理”。毋庸置疑，这种反差表明以美国为代表的少数国际主体所推行的竞争中立是经过别番精心设计的。根据以往的经验来看，这其中定然不同程度地充斥着部分混淆视听的“歪理邪说”。因此，摧邪辅正厘清竞争中立的本义显得非常必要。

（一）罗伯特·霍马茨之谬

根据罗伯特·霍马茨的解释，美国所推行的竞争中立要义是使竞争不受外来因素的干扰。按照正常思维逻辑的理解，罗伯特·霍马茨所指的“外来因素”应当为一国的政府行为；换言之，罗伯特·霍马茨所谓的竞争中立是指竞争不受一国政府行为的干扰。

众所周知，“市场是资源配置的有效机制，这已为十九世纪欧洲和北美洲大多数国家的资本主义经济成就所证明。但是，市场不是万能的，市场机制也有固有许多自身不可克服的缺陷，以致在资源配置上失灵和失效，即不能或难以实现资源的高效配置，这已为十九世纪末以来资本主义经济的发展状况，特别是周期性经济危机所证明。”[13]理论研究和历史实践充分表明，市场失灵的表现形式是有很多种的，但首当其中的便是竞争导致垄断。“列宁曾经指出：竞争必然引起资本的积聚和生产的集中，而这种积聚和集中发展到一定阶段就必然走向垄断。”[14]竞争导致垄断的市场失灵客观上决定了政府干预的必然性，政府在市场失灵范围内干预市场竞争已经成为经济学、经济法学领域的一种普世认知。但是按照罗伯特·霍马茨所言的竞争中立，为了保证国有企业和私营企业公平竞争，各国政府都应当不采取任何措施干预相关市场的竞争。很显然，这种主张无论是在理论上还是现实中都是无法立足的，且颇有因噎废食之嫌。

罗伯特·霍马茨的这番论调不免使我们情不自禁地将之与竭力倡导自由竞争的古典经济学派联系起来。然而相较而言，前者比后者更为激进。虽然以亚当·斯密为代表的古典经济学派主张自由竞争，但是亚当·斯密并不是主张经济上的自由放任。[15]在《国民财富的性质和原因的研究》中，亚当·斯密特别强调：“商人的利益在某些方面往往和公众的利益不同，有时甚至相反。扩张市场，减少竞争，这无疑是一般商人的利益。前者虽然往往于公众有利，后者却总是和公众利益相反。减少竞争，只会使得商人的利益提高到自然的程度之上，其他人却为了他们的利益而承受不合的负担。”[16]但是罗伯特·霍马茨的这番激进在性质上并不是一种无知的表现，恰恰相反，它深刻地反映出美国试图借助竞

[12] 王婷：《竞争中立：国际贸易与投资规则的新焦点》，《国际经济合作》2012 年第 9 期。
[13] 王全兴著：《经济法基础理论专题研究》，中国检察出版社 2002 年版，第 80 页。
[14] 转引自钟明钊主编：《竞争法》，法律出版社 2008 年版，第 9 页。
[15] 王晓晔著：《反垄断法》，法律出版社 2011 年版，第 9 页。
[16] [英]亚当·斯密著：《国民财富的性质和原因的研究》（下册），郭大力、王亚南译，商务印书馆 1983 年版，第 242—243 页。

争中立的字面意思所极力掩藏的内心真实想法，因为纯粹的自由竞争将导致美国企业垄断或者继续垄断全球的主要核心产品、技术、服务等市场。

（二）澳式竞争中立的误传

因上个世纪末澳大利亚境内的市场不公平竞争（垄断）问题主要系由国有企业可以通过政府减免税收、获得补贴等原因造成的，[17] 所以澳大利亚将竞争中立问题定格在政府商业活动（国有企业）和私营企业之间。假若不考虑澳大利亚的特定历史背景、具体改革目的等因素孤立地审视《联邦竞争中立政策声明》所阐述的竞争中立，它在直观上显得比较狭隘。纵览各国市场经济发展史，竞争关系盘根错节而非整齐划一的存在于不同类型的市场参与者之间，商品供给替代不仅发生在一国境内的国有企业和私营企业之间，而且发生在一国境内的国有企业之间和私营企业之间。尽管因特殊的“血缘”关系使得国有企业常常受到政府不合理的特殊关照而获得额外竞争优势，但是因政府投资的主体多元化和产权的持有多样化，这种额外竞争优势并非普遍、均衡惠及所有国有企业；况且根据实践来看，从政府不当干预市场行为中获得额外竞争优势并非国有企业的“专利”，私营企业通过商业贿赂、院外游说、政府非对称管制等获得额外竞争优势在现实生活中也是屡见不鲜。因此，竞争中立的周全之义至少应当涵盖所有企业之间的竞争。

但是澳大利亚《联邦竞争中立政策声明》文本表意的竞争中立却出人意料地格外受到美国的“青睐”，它不仅完全照搬其义塞入 TPP 协议使之走向区域化，而且借道利用 OECD 发布《澳大利亚的竞争中立与国有企业：实务和其他相关的评论》（Competitive Neutrality and State-Owned Enterprises in Australia: Review of Practices and their Relevance for Other Countries）等工作报告使之走向全球化。美国的这些断章取义之举导致了澳大利亚的竞争中立被不断的误传，不仅使得国际社会片面地将竞争中立问题狭隘地聚焦在国有企业与私营企业之间甚至仅限在中国的国有企业与其他的私营企业之间，而且使得国际社会容易忽视竞争中立的从属性而错误地将之绝对化。

不仅澳大利亚竞争中立的文本表意是有着特定的历史背景，而且澳大利亚竞争中立的生成运作也是有着特定的社会环境。根据澳大利亚的实践来看，竞争中立并非独尊生成孤立运作的，它是服务于以维护市场有效竞争为目标的公平竞争机制；竞争中立并非意味着绝对禁止政府采取任何非中立性的干预措施（包括有意识的重点扶持国有企业），《2010 竞争与消费者法》就明确允许澳大利亚竞争与消费者委员会在特定情况下采取非中立性的规制措施。因此，竞争中立具有相对性，其对政府行为的约束内容皆取决于以维护市场有效竞争为目标的公平竞争机制的场景需求。这客观上就决定了国内市场的竞争中立操作方式未必可以直接“移植”到国际市场，而美国在传播澳大利亚竞争中立的过程中似乎有意忽视了此点，这容易造成人们误以为竞争中立不仅在约束效力上具有绝对性，而且在操作方式上具有绝对性。

（三）竞争中立的含义界定

作为市场竞争法治一种理念，竞争中立应然之义可以述为：在市场有效竞争的环境下，除基于克服潜在的市场失灵需要而依法进行非中立性的干预之外，政府应当公平地对待参与市场竞争的所有企业。

[17] See Tony Greenwood, David Williamson, Jim Armitage, Gary Rumble and Donald Magarey，Corporatization and privatization of State-owned Enterprises：Some Australia Perspective，17 Hastings Int'l& Com p[J].Law Review,741,1994.

罗伯特·霍马茨之谬的驳斥表明，因竞争导致垄断引发的市场失灵客观决定了政府干预的必要性，但是“我们说政府有时可以改善市场结果并不意味着它总能这样。”[18]“当政府政策或集体行动所采取的手段不能改善经济效率或道德上可接受的收入分配时，政府失灵便产生了。”[19]政府失灵催生了多种限制政府干预的机制，发源于澳大利亚的竞争中立便是这类机制的一种。市场与政府的耦合路径表明，作为限制政府干预机制的竞争中立应当有着特定的存立条件，即市场没有出现失灵，精确而言便是相关市场存在有效的竞争。若相关市场没有竞争，亦无政府竞争中立之说。值得特别指出的是，经济学评估市场竞争状况的传统方法都是以一个抽象的市场为基础的，基本上没有虑及商品因素之外的政治问题；但是根据实践来看，国际市场的商品供给还受到政治制裁、经济制裁、出口管制等各种政治因素影响。因此，一国在评估国际市场的竞争状况过程中不仅应当重点考虑厂商数量、商品价格、产品质量等因素，而且应当合理考虑厂商国籍、本国政治影响力、本土企业参与程度等因素。

澳式竞争中立的误传之澄清不仅深度还原了政府干预所应当指向的竞争范围，而且初步指明了竞争中立的相对性。毫无疑问，在市场有效竞争的环境下，政府原则上应当保持中立公平对待所有参与市场竞争的企业；但是这种中立不是绝对的，各国在禁止政府滥用权力限制、排除竞争之际都不约而同地授权政府依法采取非中立性的干预措施以防潜在的市场失灵从而确保市场有效竞争的持续性。就以澳大利亚为例，《2010 竞争与消费者法》在第 46 节、第 46A 节、第 46B 节对具有市场支配地位的企业强加了一些特殊的法律义务（禁止滥用市场支配地位）；根据第 76（1A）条规定，违法者将可能被处以 1000 万澳元的罚金，若有违法所得，将可能被处以三倍违法所得的罚金，反之将可能被处上一年度营业额 10%的罚金。类似澳大利亚的上述规定在美国、欧盟、日本、巴西、南非等各自的竞争法律规范中都是普遍常见的，它在性质上是一种非中立性的规制措施。如德国法学家狄特瑞希·霍夫曼（Dietrich Hoffmann）形象所言，“一种行为若由其他企业实施则可能是正常的竞争，但若由拥有市场支配地位的企业实施就构成‘滥用’并受到禁止。”[20]

厘清竞争中立的本义不仅有助于我们在国际贸易谈判中有理有据地采取各种举措进行博弈以在最大程度上抵御中国发展所面临的“外患”，而且有助于我们在国内经济改革中对症下药地采取各种手段进行破局以在较短期限内消除中国发展所存有的“内忧”。

三、竞争中立的基本行为准则

在竞争中立的本义之下，无论是基于内在（国内市场）的固本培元需要还是基于外在（国际市场）的遵规守道需要，一国政府都应当在交易机会、经营负担、投资回报三个竞争点面上保持中立，确保给参与竞争的所有企业提供一个公平的市场环境。

（一）交易机会中立

交易机会中立，即政府在干预市场过程中应当公平地向所有企业提供交易机会。如诺贝尔经济学奖得主哈耶克所指出，“市场上各方必须应该自由地按照他们能找到的交易伙伴的价格进行买卖，任何人必须应该自由地生产、出售和买进任何有可能生产和出售的东

[18] [美] 曼昆著：《经济学原理》（第 3 版），梁小民译，机械工业出版社 2003 年版，第 10 页。
[19] [美]保罗·A·萨缪尔森、威廉·D·诺得豪斯著：《经济学》（下），中国发展出版社 1992 年版，第 1189 页。
[20] 转引自曹士兵著：《反垄断法研究》，法律出版社 1996 年版，第 140 页。

西。进入各种贸易的通道也必须在平等的条件下向所有人开放，法律必须不能容忍任何个人或者集团企图通过公开或者隐秘的力量限制这些通道。”[21]因此，政府在管理市场过程中应当在最大程度上确保各种贸易“通道”的顺畅性和开放性，使得各个企业都能公平地获得各种交易机会。

交易机会中立首先要求政府在进入管制上保持中立。虽然市场上的交易机会除政府采购之外本身都系由企业相互之间提供的，但是这些交易机会的获取（分配）在不同程度上却直接受到政府的进入管制影响。以斯蒂格勒列举的“管制当局对企业进入特定行业的限制，比如设立银行必须经过审批等特别程序、民用航空委员会对开通信航线的批准”为例，[22]如果申请者的请求被管制部门否决，则意味着这些企业至少在获得市场准入之前都将无法触及被管制市场的各种潜在交易机会。因此，政府在进入管制上必须保持高度的中立。“管制中立，要求最大程度上保持政府商业活动和私人企业享有同样的管制环境。扭曲的情形包括，在某些国家的某些领域，国有企业更容易取得建筑许可，政府控制的金融活动所受到的管制更少等。进一步的考虑还应该包括竞争法和反垄断条例对于国有企业和其他形式政府商业活动的适用范围。”[23]

交易机会中立其次要求政府在政府采购上保持中立。政府采购是政府作为交易相对人直接向企业提供的市场交易机会，它是除政府滥用权力排除、限制竞争之外政府有权直接决定市场交易机会获得的唯一合法场景。因政府采购的经济规模相对庞大且比较稳定，这使得政府采购对相关市场的企业竞争影响极其深远，甚至在有些情况下决定着一些企业的生死存亡。故几乎所有企业都不敢小觑任何一级政府的任何一次采购所蕴含的各种商机，政府采购的市场竞争也就格外激烈。“政府采购是当今世界各国政府管理社会经济生活的一种重要手段，既不同于一般的私人、企业采购，也不同于政府的一般行政行为。”[24]因此，作为政府采购交易相对人的政府不能像传统民商事主体那样可以“肆意”选择合作者，它必须在政府采购上保持高度的中立，这就要求政府在各种采购项目对外有效开放的条件下必须以非歧视性的标准从国有企业、私营企业中遴选合适的经营者进行缔约。

（二）经营负担中立

经营负担中立，即政府在干预市场过程中应当公平地处理各类企业的经营负担。任何企业在市场化经营过程中都面临着各种负担，既有来自政府方面单边课征的负担（强制性负担），又有来自市场方面合意课加的负担（协商性负担）。但不论这些负担来自何处，它们都直接或者间接决定着企业的经营成本，而这将直接影响着一个企业能否成功实施成本领先战略（美国哈佛商学院著名的战略管理学家迈克尔·波特提出的三大基本竞争战略之一）以获得市场竞争优势。根据各国的实践来看，政府不仅可以直接影响政府方面单边课征的负担，而且可以直接影响市场方面合意课加的负担。因此，政府在管理市场过程中应当在最大程度上一视同仁地影响所有企业可能承受的经营负担以确保市场竞争的公平性。

经营负担中立首先要求政府在对企业课征强制性负担上保持中立。企业在任何交易机会下都将不可避免地面临着政府在雇员福利、产品质量、环境保护、税收等多个方面课征的强制性负担，这类经营负担构成企业产品刚性成本的重要组成部分，它们对企业竞争力

[21] [英]弗里德里希•奥古斯特•哈耶克著：《通往奴役之路》，王明毅等译，中国社会科学出版社 1997 年版，第 41 页。
[22] [美] G. J. 斯蒂格勒著：《产业组织和政府管制》，潘振民译，上海三联书店 1989 年版，第 212—215 页。
[23] 唐宜红、姚曦：《竞争中立: 国际市场新规则》，《国际贸易》2013 年第 3 期。
[24] 王小能：《政府采购法律制度初探》，《法学研究》2000 年第 1 期。

的影响较大，特别是以货币形式直接计入企业产品成本的税收。尽管随着税收法治的日益完善，一国内的企业之间特别是国有企业与私营企业之间的税负差别基本上没有什么差异，但是在实践中客观上还存在一些扭曲市场竞争机制的税负不公平问题，如在一些尚未改制以公司形式进行商业活动的政府经营性行为上不同程度的享受减免间接税、在局部范围内外资企业比内资企业享受更多的税收优惠等；政府应当积极进一步采取措施消除这些税负不公残余问题，最大程度上保持税收中立。

经营负担中立其次要求政府在对企业消减协商性负担上保持中立。除局部私法公法化领域之外，企业在任何交易机会下的协商性负担通常都是由合同当事人合意确定的，如原料交易价格、外包服务责任、商业贷款利率等，政府原则上不应当也不会进行干预；但是在实践中客观上还频频出现政府“越位”、“错位”现象，如地方政府为本地的纳税大户企业（国企私企皆有）进行信用贷款担保、中央政府职能部门隐蔽性地指令国有控股的金融机构向央企提供低于市场利率的商业贷款等。这些现象在不同程度上影响了相关市场的公平竞争，因此，政府应当在最大程度上确保自身始终处于“正位”以保持在企业信贷等方面中立。

（三）投资回报中立

投资回报中立，即政府在干预市场过程中应当公平地影响企业的投资回报。如史学家司马迁在《史记》“货殖列传”篇章中所曰：“天下熙熙，皆为利来；天下壤壤，皆为利往。”商业化运作的企业争取交易机会、承受经营负担的一切作为都是为了追逐投资回报，而能否获得投资回报以及回报率的高低直接决定着一个企业在市场竞争中的命运。在市场经济条件下，企业的投资回报主要取决于各自的技术革新速度、成本管控绩效、市场营销能力等内在因素，但是在少数情况下还会不同程度的受到政府干预这个外在因素的影响。因此，政府在管理市场过程中应当在最大程度上保持投资回报中立以使得企业之间公平竞争。

投资回报中立首先要求政府在价格规制上中立。虽然产品定价基本已经普遍实现企业自主化，但是各国都不同程度的保留了部分干预权。除在非竞争性市场上仍然大量存在的政府定价、政府指导价等传统价格管制外，大多数国家都在竞争性市场上进行了超高定价的反垄断规制，如美国、欧盟、日本、澳大利亚、巴西、南非等。超高定价的反垄断规制如欧盟法院在General Motors Continental NV案件判决中所言，“一个占有市场支配地位的企业，如果索要的价格与其他产品或者服务的经济价值相比过高，那就基本可以认定该企业存在超高定价问题。”[25]由于企业的投资回报通常都与产品的价格高低有着紧密的正比关系，所以政府的价格规制行为在一定程度上直接影响着企业获得的投资回报；因此，政府应当在价格规制上保持中立。价格规制中立不仅要求政府尽量避免进行选择性执法，而且要求政府尽量择用科学的方法进行执法，如根据产品周期理论进行分析产品利润较高是否具有合理性等，而不能简单地以价格绝对值、行业平均利润率、消费者的期望值等指标进行武断。

投资回报中立其次要求政府在政府补贴上中立。政府补贴是企业除从市场之外获得投资回报的另一个重要途径，能否获得补贴、获得补贴额度的大小以及获得补贴的行业份额比等都在不同程度上影响着一个企业的生存与发展，这在整体经济低迷、行业产能升级、个体财政危机等特殊情况下表现的尤为突出。因此，政府应当在政府补贴上保持中立。“政府资金应减少对企业点对点的支持，加大普惠性政策支持的力度，使政府的支持有利

[25] Valentine Korah, Case and Materials on EC Competition Law, Oxford 1997, p.111

于建立良性的企业创新机制；政府应创造公平的市场环境，尊重企业的首创精神，尊重市场对产业发展方向和技术路线的筛选。”[26]

竞争中立的基本行为准则：一方面为我国现阶段深化经济体制改革的“顶层设计”提供了规划参考，政府应当积极响应商界在《中国企业发展环境报告2013》中表达的诉求为他们创造一个公平的市场环境；另一方面为我国中长期推动国际规则演化的“全球治理”呈贡了路径选择，政府应当积极借鉴美国在TPP谈判中采取的做法为中国谋求一个利好的经贸规则。

四、竞争中立的适用除外情形

竞争中立的本义表明，竞争中立在适用上并非像美国所倡导的那样似乎具有普遍性、无条件性和绝对性，那些特殊性、有条件性和相对性项下的内容构成了竞争中立的适用除外情形，政府的相应作为不受竞争中立的基本行为准则约束。

（一）非市场化领域的政府管理

竞争中立本义中的“在市场有效竞争的环境下”暗示了竞争中立只适用于市场化领域，因此，非市场化领域的政府管理行为也就不在竞争中立的“管辖”范围内。根据实践来看，非市场化领域在整体上可以分为两大个类型，即绝对的非市场化领域和相对的非市场化领域。绝对的非市场化领域是指相应资源的配置直接由政府以公权力作出命令的方式进行统筹安排的领域。随着“小政府大社会”治国理念的不断深化，绝对的非市场化领域在推行市场经济的国家日常生活中越来越少，目前主要在国家安全行业、自然垄断行业等少数产业的局部领域保有。以国家安全行业中核心军工技术研发为例，包括法国、俄罗斯、印度、阿根廷、印度尼西亚、乌克兰等在内的不少国家目前都还是采用“军队系统内部设研发机构+从军费预算中直接拨付研发费用”的模式，这部分军费的使用均以非政府采购的方式高度机密化进行。无论是对于内资还是外资或者国企还是私企，政府在这些特定领域内都无需向任何企业作出竞争中立的承诺。

相对的非市场化领域是指相应资源的配置在特殊情况下政府基于一些事由的考量而临时性、个案性或者永久性、普遍性地不向特定企业开放的领域。虽然伴随着WTO规则适用的不断拓土，国民待遇原则、无歧视性待遇原则等使得相对的非市场领域在推行市场经济的国家存在的空间日渐萎缩，但是这种现象似乎在有些国家表现的愈久愈坚。以竭力倡导竞争中立的美国为例，美国不仅从上个世纪50年代就开始发布《1950年国防生产法》（The Defense Production Act of 1950）、《1979年出口管理法》（Export Administration Act of 1979）等推动对外资的国家安全审查，而且在宣布加入TPP谈判前夕专门出台了《2007 年外商投资与国家安全法案》（Foreign Investment and National Security Act of 2007）强化对外资的国家安全审查。“外资并购交易如果威胁到美国在关键技术领域的世界领先地位，或影响美国的本土就业，都将被视为威胁国家安全。由此可见，美国为其行政审批设定了更为宽泛的范围，也让外国投资者赴美并购的风险大大增加。”[27] 中国海洋石油有限公司收购美国尤尼科石油公司的失败、[28] 迪拜世界港口公司

[26] 唐福勇：《公平的市场环境比政府补贴更重要》，《中国经济时报》2013年8月27日。

[27] 王小琼、何焰：《美国外资并购国家安全审查立法的新发展及其启示——兼论《中华人民共和国反垄断法》第31条的实施》，《法商研究》2008年第6期。

[28] See Joshua W. Casselman, China’s Latest Threat to the United States : The Failed CNOOC - Unocal Merger and Its Implications for Exon-Florio and CFIUS , 17 Ind. Int’l & Comp. L. Rev. 155 (2007) .

(DPW)收购英国半岛—东方航运公司(P&O)的流产[29]等都是这类风险导致的结果。目前，越来越多国家开始效仿美国对外资进行国家安全审查；至少按照美国之意而言，一国政府依法作出的相应裁决相对于外资而言不受竞争中立的管束。

（二）引入竞争机制的政府改革

竞争中立本义中的“在市场有效竞争的环境下”直接揭示了竞争中立在市场化领域的适用前提条件，如果一个市场化的领域缺少有效竞争，那么政府在这个领域引入竞争机制（创造竞争）的改革行为理应不受竞争中立的评价。

放手市场、引入竞争是很多国家在国内经济体制改革中主要采取的操作步调，从实践来看，政府在此过程中通常都会不同程度地采取一些非中立性的扶持措施。以英国电信行业的早期改革为例：1981 年，为了推动电信行业走向市场化，英国政府批准了成立莫克瑞通信公司(Mercury Communications)，使之与英国电信公司（BT）在国内展开竞争；“英国政府对新成立的 Mercury 公司有一定的扶持，但并不刻意削减 BT 公司的实力。”[30] “1983 年，英国政府明文规定国内只允许 BT 与莫克瑞通信公司建设经营固定网络和经营基本电信业务，从此英国的电信产业形成了双寡头垄断格局，这个格局一直维持到 1991 年，其他运营商才得以进入英国电信产业的固定网络市场。英国政府采取这一政策的目的，是通过在一定时期内建立和维持一个双寡头垄断的格局，帮助新成立的莫克瑞公司建立起自己的通信网络，使其不断地扩大市场份额，增强与 BT 的竞争能力，适应未来激烈的市场竞争。”[31]针对这些引入竞争机制的政府改革行为，因其道义不仅与竞争中立同向，而且视野高于竞争中立，所以应当不受竞争中立的羁束。

如前面所言，由于国际市场的商品供给可能不同程度的受到政治制裁、经济制裁、出口管制等政治因素的影响，所以各国政府在评估国际市场的竞争状况过程中除了需要考虑厂商数量等市场核心因素之外，还必须合理考虑厂商国籍等相关政治因素。如果一国的特定物资供给基本由外国企业实际掌控并且较为容易受到国际政治因素的影响，即便厂商数量并非呈现独占或者寡头特质，该国政府也可以认定此市场化领域（相对）缺乏有效竞争。以航空发动机为例，虽然通用电气公司、普拉特·惠特尼集团公司、罗尔斯·罗伊斯公司、斯奈克玛公司、雷宾斯克乌法发动机生产联合体和彼尔姆航空发动机股份公司、莫斯科克里莫夫公司、国际航空发动机公司、CFM 国际发动机公司等企业都向国际市场供给飞机发动机，但是这些企业几乎都是美俄英法德的本土企业且它们的产品出口都很容受到本国政府的政治外交影响，所以相对于主要依赖于进口的印度、澳大利亚以及我国等国家而言，这类国际市场可以视为垄断市场。在保持不采取任何措施刻意打击其他市场参与者的前提条件下，一国政府在此情形下有权通过产业政策在人力、资金、技术等方面倾斜性地培育和扶持本土企业，包括以国家资本形式存在的国有企业和以民间资本形式存在的私营企业。

（三）维持竞争机制的政府规制

竞争中立本义中的“除基于克服潜在的市场失灵需要而依法进行非中立性的干预之外”突出了竞争中立在市场化领域现行处于有效竞争情况下的非绝对性，政府为了使得这种有效竞争格局能够动态性的持续下去（维持竞争）而依法采取非中立性的干预措施（政

[29] See Deborah M. Mostaghel, Dubai Port s World Under Exon-Florio : A Threat to National Security or a Tempest in a Seaport ?, 70 Alb. L. Rev. 583 (2007) .

[30] 袁正、高伟：《中国电信业改革回顾、经验与问题》，《宏观经济研究》2009 年第 9 期。

[31] 刘戒骄：《英国电信产业的放松管制和对主导运营商 BT 的再管制》，《中国工业经济》2002 年第 1 期。

府规制）亦不受竞争中立的辖制。熊彼特的创造性破坏理论(Schumpeter's creative destruction) 不仅揭示了市场竞争的动态特质，即产业领先者与挑战者攻击与反击的市场过程；而且揭示了除政府滥用权力限制、排除竞争之外可以妨碍市场有效竞争的另一个因素，即企业（创新的先动者）拖延报复（竞争对手的反击）或使时滞（竞争对手的反应时间）最大化的战略行动。因此，市场有效竞争的维持不仅需要推行竞争中立严格约束政府干预市场竞争的行为，而且需要实施政府规制严格约束企业影响市场竞争的行为。由此可见，政府规制与竞争中立在维护竞争机制的保障体系中是具有同等相称地位的，所以前者理应不受后者的辖制。

政府规制的操作方式在大多数情况下是给部分企业强加一些特殊的法律义务，这主要表现在各国的竞争法律规范上。澳大利亚、美国、欧盟、日本、巴西、南非等在各自的竞争法律规范中除了规定禁止滥用市场支配地位（包含禁止知识产权滥用）外，还规定了经营者集中的竞争审查。“‘经营者集中’是借鉴了欧共体竞争法的概念‘Concentration between Undertakings’。美国和其他很多国家的反垄断法不使用‘经营者集中’，而是使用‘企业并购’，即 Merger & Acquisition ，简称 M&A。”[32] 竞争审查并非针对所有的企业并购，鉴于哈佛学派曾经奉行的结构主义所基于的现代产业经济学“结构—行为—绩效”分析范式（Structure-Conduct-Performance）具有的科学性一面，各国政府通常只对达到法定申报标准的企业并购根据“市场支配地位标准”或者“实质性减少竞争标准”进行竞争评估并据此相应作出禁止性裁决、无条件批准裁决或者附加限制性条件裁决。在少数情况下，政府规制的操作方式是给一些企业提供部分的特殊好处，如授予某种特权、给予特定资助等；前者如各国竞争法律规范中对中小企业为了提高竞争力而实施的横向垄断协议进行豁免情形，后者如很多国家对科技孵化园区的企业给予数额不等的科研经费支持。

竞争中立的适用除外情形为竞争中立的适用范围划定了疆界，这是竞争中立的客观规律固有内容，任何国家政府都应当遵循该市场规律。因此，我们在坚持以竞争中立来防治政府滥用权力限制、排除竞争之时应当充分尊重政府在竞争中立的适用除外情形下所实施的正当作为，并且各国政府应当相互尊重彼此遵循上述市场规律的各自做法。

五、竞争中立问题的治理路径

受权力寻租、贸易保护主义、竞争意识淡薄等诸多方面因素的影响，（各国）政府在干预市场过程中都会不同程度的出现滥用权力限制、排除竞争问题。因此，如何有效使得政府保持竞争中立仍然是个世界性的社会治理难题。竞争中立问题的复杂性客观上决定了竞争中立问题的治理必须走一条多层多维制约机制齐抓共管之路。

（一）竞争法律规制

一国对本国政府的竞争中立的内在刚性约束在正常情况下应当主要依赖于竞争法律规制，即通过竞争立法禁止政府滥用权力限制、排除竞争。因参加主体、交易行为的类型、数量等非常庞杂，所以市场经济的有效运行必须得到各种具有约束力规则的全面支撑，故市场经济必须是法治经济。[33] 由于竞争是市场经济的灵魂，因此，法律应当禁止包括政府在内的任何主体在市场没有出现失灵的情况下人为地限制、排除竞争，任何竞争违规行

[32] 王晓晔：《〈中华人民共和国反垄断法〉析评》，《法学研究》2008 年第 4 期。
[33] 参见刘武俊：《深刻理解市场经济是法治经济》，《人民日报》2012 年 7 月 2 日。

为都必须及时有效的被竞争法律所规制。诚如澳大利亚的实践所示，尽管治理政府滥用权力限制、排除竞争所产生的竞争中立问题是一项较为系统性的社会工程，它不仅涉及到竞争法律的实施问题，而且涉及到政治、经济甚至文化等问题；然而，一切治理工作的开展都是以《2010 竞争与消费者法》等竞争法律为基本准绳的。市场经济的法治要求和政府滥用权力限制、排除竞争的要件特性客观上决定了竞争法律在治理竞争中立问题中的基础地位与决定性作用，竞争法律规制是治理竞争中立问题的基本、主要路径。

事实充分证明了这点，市场经济和法治水平相对发达的美国、欧盟、日本、韩国及其澳大利亚等所采取的各种应对措施都是围绕着各自的竞争法律进行的，他们只是在模式与手段的直接性/间接性程度上有所差异而已。受政治体制、经济模式、立法博弈等因素的影响，各个国家的竞争法律所禁止的政府滥用权力限制、排除竞争中的抽象“政府”在具体范畴上并非完全一致的。多数一般仅局限于行政机构，即禁止政府滥用行政权力限制、排斥竞争（通称行政性垄断），如美国的反托拉斯法律制度、日本的《禁止私人垄断及确保公正交易法》、韩国的《规制垄断与公平交易法》以及澳大利亚的《2010 竞争与消费者法》等；少数涵盖立法机构、行政机构、司法机构，典型代表就是欧盟的《欧共体条约》。竞争中立所规制的抽象“政府”具体范畴越广，竞争中立问题的治理归口至少在形式上就越简化（集中管治模式），在正常情况下由竞争执法机构依法采取立法优先咨询等事前预防措施、承诺机制等事中化解措施、责任人刑罚等事后矫正措施进行即可；反之，竞争中立问题的治理归口至少在形式上更为复杂（分散管治模式），通常还需要像澳大利亚那样安排其他部门介入并在竞争执法之外进行其他配套的政治协商、经济改革等。

虽然《中华人民共和国反垄断法》对政府滥用（行政）权力限制、排除竞争仅作了宣示性的规定，这注定了我国在现行条件下对竞争中立问题的治理基本依赖于政府的自我约束和行政法律的偶合管控；但是这些宣示性规定所起到的警告性和依据性使得竞争法律在中国竞争中立问题的治理中的地位和作用并未受到任何影响，只是在模式与手段上带有竞争执法色彩的内容相对其他国家或者地区少得可怜，只有“反垄断执法机构可以向有关上级机关提出依法处理的建议”。为了更好地发挥《中华人民共和国反垄断法》对竞争中立问题的规制功效，我国应当通过修改立法不仅赋予反垄断执法机构对行政性垄断的刚性规制权，而且授予反垄断执法机构更多相应的处置手段。

（二）国贸协定约束

一国对本国政府的竞争中立的外在刚性约束原则上只有国贸协定约束，即通过缔结双边或者多边国际贸易协定制约各方的政府滥用权力限制、排除竞争。虽然几乎各个国家和地区的竞争法律都作了类似《中华人民共和国反垄断法》第二条域外适用制度规定，但是因国家主权之间的平等性使得这些规定只适用于经营者的垄断行为而不适合他国政府滥用权力限制、排除竞争。因此，一国如果希望能够有效制约他国政府滥用权力限制、排除竞争，在正常情况下唯一可行之计就是与他国签订平等适用于各方的国际贸易协定；换言之，一国政府在竞争中立问题上的外来刚性约束只有其与他国自愿签订的国际贸易协定。

虽然目前缔约国家数量最多的 WTO 的现行贸易规则对较为直接影响企业在国际市场上公平竞争的关税、倾销、补贴等事项作了刚性约束，但是不仅这些内容远远窄于竞争中立本义之下的交易机会中立、经营负担中立与投资回报中立的各个方面要求，而且势头再起的多国双边贸易协定谈判与各式各样的国际区域性自贸区谈判使得 WTO 本身的发展前景非常堪忧。这客观上恶性循环地导致了国际市场竞争中立问题的规制必然走向游离于现行 WTO 框架之外其他新的国际贸易规则，现实已经充分表明了这种发展趋势：不仅作为 WTO

重要创始国之一的美国已经另起炉灶在大力推行渐成气候的 TPP，而且很多权威专家纷纷主张着手制定一部全球性的竞争法律。例如，美国著名的国际法专家戴维·格伯尔教授指出："如果全球竞争缺乏一个有效的法律框架，反竞争行为就会损害市场效率，进而使各国的人民无法享有全球市场所产生的经济资源和机会，那些需要最为迫切的人们尤其被剥削了上述好处。因此，在全球市场上建立一个更为有效的法律体系应对反竞争行为是非常必要的"。[34]

尽管 OECD 在向各国推广《竞争法基本框架》（A Framework for Competition Law）过程中所遇到的困难表明在全球施行一部统一的竞争法律在短期内是不切实际的，但是欧洲竞争法律统一的成功示例和 TPP 历史演进说明在特定国际区域内制定和实施一部统一的竞争法律在预期时间内还是可行的。因此，国际市场上竞争中立问题的规制在未来数年内必将呈现出国际区域化格局，而这种格局将给中国企业"走出去"的发展战略带来巨大的影响。[35] 如果我国不能及时成功地融入到这些国际区域化的经济体中去，中国企业将直接面临着不同程度被边缘化的风险。所以我国不仅应当及早参与诸如 TPP 之类新的国际贸易协定谈判，而且应当主动出击打造多个经济版本的"上合组织"。庞大的经济体量、优良的企业质量与丰富的市场经验已经使得我国有着深厚的社会资本和强大的综合能力来承受这些新的国际贸易协定可能带来的各种阶段性发展压力。

（三）竞争文化引导

竞争文化，概括地说就是关于市场竞争的一系列思想观念、商业规则和法律制度的总称。[36] 它对于引导包括各国政府在内的所有市场参与者切实遵守竞争法律和国贸协定是有着非常重要的现实意义，起到软性补充作用。

优胜劣汰法则牵扯的伦理问题使得人们原本对竞争就存在认识分歧，如亚当·斯密认为"卓越源自竞争和超越"，而托马斯·霍布斯却觉得"竞争使人类堕落为贪婪的猛兽"；[37] 经济利益的深度卷入更是使得人们对竞争的想法复杂多变，即便被公认利益高度超脱化的政府也会时不时的出现迷失错位情形。"竞争主管机构以外的其他机构及大众缺乏对竞争及竞争法原理的认同，也缺少维护、促进竞争的内在责任感。即使企业效率低下或者存在限制竞争行为，但在涉及本地区经济发展时，地方政府常常睁一只眼闭一只眼，甚至为其提供便利。"[38] 因此，"现实中的某些法律、政策不但没有促进竞争，相反导致了对竞争的损害。"[39] 所以各国在加强竞争立法与执法力度同时，还应当加大竞争倡导（亦称竞争推进）工作，全面推动本国竞争文化的发展。深厚的竞争文化不仅使得包括政府在内的所有市场参与者容易对竞争法律产生敬畏而被迫遵守，而且使得包括政府在内的所有市场参与者容易对竞争法律产生信仰而主动维护。

受传统"和气生财"思想、市场经济施行历史较短等因素影响，"与反垄断法实施的其他问题相比，中国市场环境中竞争文化缺失的问题尤为严重。"[40] 竞争文化的严重缺

[34] [美]戴维·格伯尔著：《全球竞争：法律、市场与全球化》，陈若鸿译，中国法制出版社 2012 年版，第 6 页。
[35] 参见何曼青、柴林涛：《中国科技型企业"走出去"：成效及政策建议》，《国际经济合作》2012 年第 5 期。
[36] 徐士英、应品广：《竞争文化的培育和发展——从日本竞争主管机关竞争执法、竞争推进谈起》，《江苏大学学报(社会科学版)》2011 年第 5 期。
[37] R. Hayward, Markus Kemmelmeier, How Competition Is Viewed Across Cultures: A Test of Four Theories,41 Cross-Cultural Research 364(2007).
[38] 张占江：《竞争倡导研究》，《法学研究》2010 年第 5 期。
[39] See F.A. Hayek, The Road to Serfdom: Text and Document, Chicago: University Chicago Press, 2007, p.87.
[40] 黄勇、江山：《反垄断法实施的文化维度论纲————以竞争文化、诉讼文化与权利文化为中心》，《江西社会科学》2008 年第 7 期。

失不仅导致我国境内企业实施垄断行为的案件频发，而且导致我国还时不时的出现行政性垄断案件。例如，江西省弋阳县政府在2013年9月26日发布了《弋阳县人民政府办公室关于进一步加强全县猪肉类产品市场监督管理工作的通知》，违反行政法律、法规在当地实行外埠冷鲜猪肉白条肉等猪肉类产品市场准入制度以阻止双汇集团进入弋阳市场。这些行政性垄断行为不仅严重损害了政府的社会形象，而且极大影响了市场的竞争秩序。因此，培育竞争文化引导包括政府在内的所有市场参与者遵守《中华人民共和国反垄断法》及其相关约束性规则成为我国迫在眉睫需要开展的重要工作。

竞争中立问题的治理路径为各国解决政府滥用权力限制、排除竞争问题提供了基本操作思路，我们应当根据自身的行政管理体制、经济发展需求以及国际关系等具体情况来进一步采取周密的方案深化政府干预市场的方式改革。

结语

研究表明：竞争中立问题是一个错综复杂的社会问题，它不仅在性质上涉及到国内与国际、法律与政治、理念与手段等问题，而且（作为竞争法治理念）在内容上涉及到基本行为准则、适用除外情形和治理路径等问题。这就意味着我们不应当像罗伯特·霍马茨那样孤立、抽象的引用竞争中立对特定经济现象进行笼统评价，而必须根据竞争中立的应然之义及其规则全貌对特定经济现象进行具体评价。

作为竞争法治理念，竞争中立应当得到各方尊重和践行。我国对此已经给予了高度重视，中国共产党的第十八届三中全会开创性地提出“发挥市场在资源配置中的决定性作用”，这是我国深化市场经济体制改革、推进竞争中立市场法治建设的重要创举。竞争是市场的灵魂，要充分发挥市场在资源配置中的决定性作用就必须切实有效地采取各种措施保护市场竞争机制。市场竞争机制的维护不仅需要政府积极作为预防和制止经营者实施垄断行为，而且需要政府严于律己禁止滥用权力限制、排除竞争。所以，李克强总理在上任之处就明确要“砍掉政府那只乱摸的手”。改革开放特别是加入WTO近十余年的实践充分表明，政府不合理的保护无法真正给企业带来强劲稳健的市场竞争力，能够真正经过市场洗礼的企业特别是国有企业方能承载中国经济发展的重任。因此，政府在推进“经济改革2.0版本”过程中应当借助上海自由贸易区的实验平台学习澳大利亚的经验大力推进竞争中立的改革，尽快消除我国目前仍然客观存在的一些政府差别对待国企与民企问题。

但是我们坚决反对和抵制任何人利用竞争中立之名行贸易保护之实。无论对于国内市场还是国际市场，竞争中立应当始终保持作为竞争法治理念的本性，不能沦为国际博弈的政治手段。竞争中立的存立之本在于对公平价值的追求，而公平对形式与实质的兼顾需求特别是对实质的强调决定了竞争中立不可能是无条件、绝对化的，这就要求各国政府必须最低限度地合理考虑相关市场中不同竞争群体之间的“先天”差异。任何有意无意忽视这些因素的竞争中立在性质上都不同程度的蜕化为既存强势地位的企业挤压竞争者的辅助手段，若这种情形发生在不同国家的政府之间特别是发达国家与发展中国家之间便成为一种带有很强政治色彩的国际贸易保护手段；而这不仅将最终导致实质性公平的价值无法实现，而且将可能导致市场趋于高度集中化甚至寡头或独占。

作者信息：丁茂中（1981-），男，江苏盐城人，竞争法学博士，上海政法学院竞争法研究中心执行主任、副教授。

联系方式：

地址：上海市外青松公路 7989 号　　邮编：201601
电话：021-39227181
手机：13564883986
电邮：dingmaozhong@126.com

服务业的国际比较--中国应向其他国家学习什么

陈岳云
洛杉矶西来大学

摘要

本文主要侧重于对不同国家的服务业分析，并通过此全面的研究和国际比较，希望像中国这样的发展中国家可以多向发达国家学习，同时更好地发展自身服务业。

经济在飞速发展的过程中，在国内生产总值和就业方面，服务业将突显出其重要地位。但制造业的快速萎缩和服务业快速扩张可能会导致严重的问题，例如，生产效率的降低及中等收入陷阱问题。

为了维持经济增长，中国需要更好地拓展和发展服务业，特别是完善和发展社会化，专业化和制造/生产服务，及高质量的服务出口。

关键词：服务业，经济结构，经济发展，与中国经济。

International Comparisons of the Service Industry—
What China Can Learn from Other Countries

Yueyunn (Bill) Chen, University of the West, Los Angeles

Abstract

This paper focuses on analyses of the service industry in different countries over the years. Through this comprehensive study and international comparisons, developing countries like China can learn from others and better develop its service industry.

When an economy advances, its service industry will be increasingly important in terms of shares of GDPs and employment. But fast-shrinking of the manufacturing industry and rapid expansion of the service industry may cause serious problems such as being weightless, lowering productivity and suffering from the mid-income trap.

To sustain its economic growth, China needs to better expand and develop its service industry, particularly improve and develop social, professional and manufacturing/production services and improve and increase its service export.

Keywords: Service Industry, Economic Structure, Economic Development, and China's Economy

I. Introduction

The development of the service industry is crucial to a country's economy. The level of a country's service industry, particularly its ratio of the service industry added-value to the gross domestic product (GDP), indicates the country's economic development stage. In other words, when a country advances from an underdeveloped to developing and then developed economy, its service industry will be increasingly important and its percentage of the GDP and percentage of employment in the service sector will rise.

Many developing economies rely on its manufacturing industry. This is particularly true for export-oriented countries. When a country is underdeveloped or is at the beginning of its development, the total demand for service is low. The priority of the economy is to feed people and provide necessary products—mainly agricultural and manufacturing products. Also, in order to maintain its rapid economic growth, such a country needs to export manufacturing products

since it does not have any competitive advantages in its service products. The fact that 70 percent of the world trade is in manufacturing products further confirms that an export-oriented developing economy needs to rely on its manufacturing industry.

After a country has experienced economic development for many years, particularly rapid growth, the country's living standard is raised significantly; then the cost of producing these manufacturing products rises very fast. As a result, their manufacturing products will be less competitive in the world market. When combined with a retrenching world economy, then this kind of manufacturing and export-oriented country will be faced with economic trouble. That is what China has experienced in recent years. In order to stabilize its economy, a country like China must grow its service industry.

China's experience is not unique. Many rapidly growing developing countries such as South Korea and Japan have had similar experiences decades ago. But expanding the service industry and over-shrinking the manufacturing industry may cause new problems to these developing countries. For example, its economy becomes weightless and less productive and competitive. Therefore, it will be valuable to conduct a comprehensive and dynamic study of the service industry's developments in different countries and provide international comparisons. From this, emerging countries like China can learn from the others' similar experiences.

Each country is different in terms of its available natural resources, population and other factors that will affect and decide its economic structure and international competitive advantages. Both Germany and the US are developed countries, but the ratio of the service GDP to its total GDP is about 70% in Germany and 80% in the US. Therefore, one can learn from other countries' relevant experiences, but each country needs to develop and implement its economic strategy based on its unique conditions. Particularly an economic strategy needs not only to be based on long-term clear objectives and goals and sound execution, but also to be stable and well balanced. Dramatic changes in economic policies and strategies will cause more troubles than the solutions.

The rest of this paper is organized as follows: Section II reviews relevant literature related to the service industry; Section III analyzes relevant data; Section IV discusses what China can learn from other countries and what it should do; and the last section concludes the paper.

II. Review of Literature on the Service Industry

An economy consists of three sectors--agriculture, industry and service. The development of the service industry will change a country's economic structure. Such a development is always directly related to a country's economic growth. Therefore, the literature on the service industry development is included in both the economic structure and economic growth. In addition, a country's development of its service industry is related to its industrialization and modernization. Thus there are studies on the service industry from perspectives of the industrialization/modernization. One can also study the service industry in terms of international trade, particularly globalization.

Economic Structure Theory

Silva and Teixeira (2008) and Kruger (2008) conducted comprehensive surveys of literature on economic structural change. The root of the study on the economic structure can be traced back to Turgot (1766) and Smith (1776). Fisher (1939, 1952) defined the economic sectors and indicated that economic structure change occurs when the consumer's preference changes and/or the productivity of the specific sector gains. Kuznet (1966) gave arguments on why a country's economic structure changes and emphasized the four factors that will lead to the structural change--increasing demand for non-agricultural products, higher incomes and demand for importing products, international trade with less developed countries, and technological progress.

Baumol (1967) raised a serious issue on economic structural change called 'cost disease.' It involves a rise of salaries in jobs that have not experienced any labor productivity increase. Although the productivity in the service is not increased, firms in this industry need to pay more in order to retain and attract workers because the productivity in other sectors (industry sector) has risen and therefore salaries rise there. The associated outcome from this cost disease is that productivity growth in the whole economy will slow down when resources shift to service industry. Oulton (2001) showed that the above outcome is true only for service industries which provide final goods but not true for the ones providing intermediate goods.

Economic Growth Theory

What is the relationship between economic growth and structure change? According to neoclassical growth theory (Solow (1956, 1957)) and new growth theory (Lucas, 1988) (Romer,1986, 1990), technological progress is the only main factor contributing to economic growth. Economic growth warrants transfer of resources from low-productivity primary production to high-productivity secondary sectors. At the mature stage, resources then are transferred from secondary to tertiary sectors (Clark 1958, Rostow 1960, Lewis 1972). That is what the world economy has generally experienced, i.e. when the economy advances, more resources will be shifted from the agricultural sector to the industry sector, and then to the service sector.

Pelka (2005) developed theoretical models and found that only the economic growth process promotes structural change. However, Dietrick (2009) used OECD data from 1960-2004 and found evidence that economic structure change causes aggregate economic growth, and conversely aggregate growth leads to economic structure change. In other words, the relationship between economic growth and structure change is dynamic and interactive.

Industrialization/Modernization and Service Industry

Industrialization accelerates technological progress and improves productivity. That further leads to rapid shifts in production and labor from primary/agriculture to the secondary/industry sector. Consequently, the service sector rises as supported arguments stated by Kuznet (1966). Hansen and Prescott (2002) showed the importance of technology to growth after industrial revolution.

Modern economic growth depends on the shift from a pre-industrial land-intensive Malthusian technology with decreasing return to labor, to a Solowian constant returns-to-scale technology, with both capital and labor as inputs. Post-industrial theory (Touraine,1969, Bell, 1973), Castells and Aoyama, 1994)) states that the more advanced an economy, the more its employment and production would be focused on services. The one main source of such shifts is the information and its technology. Knowledge and information are the main sources of productivity and growth, and so the modern economy has become the information economy (Porat, 1977, Nelson, 1981, Monk, 1989, Denison, 1985, Sautter, 1976, Baumol et al, 1989).

Castells and Aoyama (1994) conducted an empirical study of employment structure in G-7 countries from 1920-1990 and concluded that the post-industrialization led to:

"the phasing out of agricultural employment, the steady decline of manufacturing employment, the rise of both producer services and social services, the increasing diversification of service activities as sources of jobs, the rapid rise of managerial, professional and technical jobs, the formation of a white-collar proletariat of semi-skilled clerical and sales workers, a substantial and relatively stable share of employment in the retail trade, and the overall upgrading of the occupational structure over time, with an increasing share going to occupations that require higher skills and advanced education."

The authors further proposed two different informational models: (1) a "service economy model" represented by the US, UK and Canada. These countries have had rapid decline of manufacturing employment after 1970s and (2) an "info-industrial model" represented by Japan and Germany. In these two countries, the share of manufacturing employment has reduced significantly, but it has maintained at a relatively high level (about 25%). Other G-7 countries fill in between these two models.

Although the service industry becomes increasingly important in a developed economy, many services and its jobs are directly related with the manufacturing industry. Cohen and Zysman (1987) argued that the post-industrial US economy is not the service economy but is a "myth". They estimated that 24% of the US GDP comes from the value added by manufacturing firms and another 25% of GDP comes from the contribution of services directly linked to manufacturing.

There are different types of services, including finance, real estate, health, and education. Stiglitz (2012) emphasized that the US needs to develop a creative economy and transfer jobs from manufacturing to services that people want—into productive activities that increase living standards, not those that increase risk and inequality. He particularly suggested that governments should invest in education and support basic research. Potts and Mandeville (2007) emphasized the importance of innovation in services, and they argued that services firms should better use information and communication technologies to coordinate service production and delivery.

International Trade and Service Industry

As Kuznet (1966) argued, increasing income will raise demand for importing products, and also trade with less developed countries will affect a country's economic structure. As we have witnessed, developed countries produce less and less manufacturing products and they buy these products from developing or underdeveloped countries. The service industry becomes increasingly significant and important in developed countries along with rapid increase in international trade and globalization.

International trade in service has been increasing in the past decades. Some services must be localized such as haircuts and facility maintenance. But others can be traded. Financial services now has been highly internationalized and globalized. Western banks and insurers are successfully expanding their businesses in developing countries. Productivity and competitiveness are the keys to the service trade.

Globalization has made the world more flat. Besides the significant increase of the world commodity trade, globalization brings more foreign direct investments and makes financial services more important.

III. Data Analyses and Implications

1. The World Total GDP and Service GDP

Table 1. World Total GDP and Service GDP

	World Total GDP (Current US$)	% of Industry GDP	Total World Industry GDP	% of Service GDP	World Total Service GDP
1995	30,332,592,675,000	32.8	9,949,090,397,400	60.8	18,442,216,346,400
2000	32,981,132,788,000	29.3	9,663,471,906,884	66.7	21,998,415,569,596
2005	46,468,942,222,000	28.3	13,150,710,648,826	68.3	31,738,287,537,626
2006	50,334,896,708,000	28.6	14,395,780,458,488	68.2	34,328,399,554,856
2007	56,696,185,105,000	28.4	16,101,716,569,820	68.4	38,780,190,611,820
2008	62,170,569,711,000	28.1	17,469,930,088,791	68.9	42,835,522,530,879
2009	58,885,873,283,000	26.4	15,545,870,546,712	70.6	41,573,426,537,798
2010	64,552,742,595,000	27	17,429,240,500,650	70	45,186,919,816,500
2011	71,448,828,400,000	27.2	19,434,081,324,800	69.7	49,799,833,394,800
2012	72,908,414,851,000	26.9	19,612,363,594,919	70.1	51,108,798,810,551

The above table shows the world total GDP, world total service GDP, and percentage of the service GDP to its total GDP since 1995.

The world total service GDP was about 70% of the world's total GDP in 2012; it was about 67% in 2000 and about 60% in 1995. Overall, the service industry has been increasingly important.

2. The World Total Trade and Total Service Trade

Table 2. World Total Trade and Service Trade

	World Total GDP (Current US$)	% of Total Trade to GDP	World Total Trade	% of Service Trade to Total GDP	World Total Service Trade	% of Service Trade to Total Trade	Total World Export (Current US$)	Total World Service Export (Current $US)	% of Service Export to Total Export
2005	46,468,942,222,000	55.12	25,613,680,952,766	10.87	5,051,174,019,531	19.72	12,781,517,318,000	2,583,043,928,860	20.21
2006	50,334,896,708,000	57.92	29,153,972,173,274	11.29	5,682,809,838,333	19.49	14,706,110,101,000	2,922,293,893,454	19.87
2007	56,696,185,105,000	58.70	33,280,660,656,635	11.92	6,758,185,264,516	20.31	17,149,064,383,000	3,486,259,004,445	20.33
2008	62,170,569,711,000	60.83	37,818,357,555,201	12.29	7,640,763,017,482	20.20	19,650,654,712,000	3,914,341,409,167	19.92
2009	58,885,873,283,000	52.38	30,844,420,425,635	11.69	6,883,758,586,783	22.32	15,755,009,644,000	3,523,647,209,987	22.37
2010	64,552,742,595,000	57.42	37,066,184,798,049	11.66	7,526,849,786,577	20.31	18,773,432,137,000	3,852,938,794,679	20.52
2011	71,448,828,400,000	61.11	43,662,379,035,240	11.80	8,430,961,751,200	19.31	22,180,434,646,000	4,328,262,400,301	19.51
2012	72,908,414,851,000	60.66	44,226,244,448,617	11.80	8,603,192,952,418	19.45	22,433,853,716,000	4,470,515,252,879	19.93

Table 2 provides the world total trade, world total service trade, and service export as well as the relevant percentage of the service trade and service export.

The world total trade has been increasing over the years. It was about 61% of the world's total GDP in 2012 while it was about 55% in 2005. According to the World Bank data, the world's total service trade was about 20% of the total trade, and the service export was also

about 20% of the world's total export in 2012. However, based on the analysis by McKinsey (2012), the world's service export composes about 30% of the world's total export.

3. GDP per capital in China and other countries

Table 3. GDP and GDP per capital in China and Other Countries (Data from IMF)

	China		US		Germany		Japan		United Kingdom		South Korea		India		Singapore		Russia		Brazil	
	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $	GDP-Bil US$	Per Capita $
1980	303.37	307.35	2862.48	12575.57	826.14	10699.04	1086.99	9311.79	542.45	9629.89	64.39	1688.84	181.42	265.91	12.05	4990.19			148.92	1256.01
1985	307.02	290.05	4346.75	18231.83	639.70	8358.02	1384.53	11464.24	468.96	8292.22	98.50	2413.94	237.62	313.07	18.46	6748.29			231.76	1742.55
1990	390.28	341.35	5979.55	23913.66	1547.03	19500.58	3103.70	25139.58	1024.56	17900.27	270.41	6307.66	326.61	385.41	38.84	12745.06			465.01	3172.10
1995	727.95	601.01	7664.05	28762.68	2525.02	30861.58	5333.93	42516.46	1181.01	20353.45	531.14	11778.76	366.60	391.25	87.06	24702.00	313.45	2113.63	769.74	4844.95
2000	1198.48	945.60	10289.73	36450.14	1891.93	22999.57	4731.20	37303.81	1496.61	25415.32	533.39	11346.66	476.64	463.12	94.31	23413.77	259.70	1775.13	644.73	3764.21
2005	2256.92	1726.05	13095.43	44224.13	2771.06	33613.84	4571.87	35780.57	2324.18	38585.28	844.87	17550.88	834.22	748.85	125.43	29403.39	763.70	5310.88	882.04	4809.85
2006	2712.92	2063.87	13857.90	46358.36	2905.45	35296.70	4356.75	34076.75	2486.60	41043.81	951.77	19676.11	949.12	839.93	146.01	33174.49	989.93	6912.93	1089.26	5869.97
2007	3494.24	2644.56	14480.35	47963.56	3328.59	40484.99	4356.35	34038.35	2858.18	46866.10	1049.24	21590.17	1238.48	1080.70	178.26	38848.26	1299.70	9101.56	1366.85	7284.37
2008	4519.95	3403.53	14720.25	48307.78	3640.73	44397.84	4849.19	37865.07	2709.57	44131.29	931.41	19028.07	1223.21	1052.67	190.32	39326.78	1660.85	11630.58	1653.54	8720.60
2009	4990.53	3739.62	14417.95	46906.90	3306.78	40424.06	5035.14	39321.22	2217.43	35885.34	834.06	16958.65	1365.34	1158.91	190.16	38127.43	1222.65	8567.94	1622.31	8472.44
2010	5930.39	4422.66	14958.30	48294.15	3310.60	40495.85	5495.39	42916.74	2296.93	36891.36	1014.89	20540.18	1708.54	1430.19	233.29	45953.53	1524.92	10671.21	2142.91	11088.60
2011	7321.99	5434.36	15533.83	49797.25	3631.44	45207.64	5905.63	46175.36	2464.64	38945.08	1114.47	22388.40	1880.10	1552.55	272.32	52533.15	1893.79	13252.56	2474.64	12693.89
2012	8229.38	6077.65	16244.58	51708.98	3427.85	42569.47	5937.77	46530.38	2484.45	38999.21	1129.60	22590.16	1858.75	1514.63	284.30	53516.04	2004.25	14015.75	2247.75	11437.39
2013	9181.38	6747.23	16799.70	53101.01	3635.96	44999.50	4901.53	38491.35	2535.76	39567.41	1221.80	24328.98	1870.65	1504.54	295.74	54775.53	2118.01	14818.64	2242.85	11310.88
% change	30.27	21.95	5.87	4.22	4.40	4.21	4.51	4.13	4.67	4.11	18.98	14.41	10.31	5.66	24.55	10.98	6.76	7.01	15.06	9.01

A country's economic development level will affect its economic structure, particularly its service sector's GDP and employment shares. Table 3 shows the total GDP and GDP per capita of selected countries from 1980-2013.

During 1980-2013, China's GDP was increased by 30.27 times and its GDP per capita by 21.95 times; US total GDP increased by 5.87 times and per capita by 4.22; Germany total GDP by 4.40 times and per capita by 4.21; Japan's increases were 4.51 and 4.13; UK was 4.67 and 4.11; Korea was 18.98 and 14.41, and Singapore was 24.55 and 10.98. At the same time, other BRIC countries experienced the following rates: India's GDP increased by 10.31 and per capita by 5.66; Russian increased by 6.76 and 7.01 (from 1995 to 2012) and Brazil increased by 15.06 and 9.01, respectively.

Based on the above statistics, China had the highest percentage increases in total GDP and GDP per capita during these years. UK had the lowest GDP per capita increase and Germany had the lowest GDP increase during the time period.

Also, among all of these countries, in 2013 Singapore and the US had the highest GDP per capita (above $50,000); Germany, UK and Japan were the next group, each with about $40,000 GDP per capita; Russia and Brazil were above $10,000 GDP per capita; China was about $7,000 GDP per capita; and India had the lowest with about $1,500 per capita.

4. China and Other Countries' Service GDP and Service Employment

Table 4 shows the percentage of service GDP to its total GDP and percentage of service employment to its total employment in China, US, UK, Japan, South Korea, Singapore, India, Russia, and Brazil.

Table 4. GDP and Employment in Service Industry

		China		US		Germany		Japan		nited Kingdo		South Korea		India		Singapore		Russia		Brazil	
		% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs	% GDP	% Jobs
1980	Agriculture	30.17	68.69	2.90	3.59	2.40	5.20	3.08	10.39	2.10	2.59	15.11	34.00	35.38		1.56	1.29			11.01	
	Industry	48.22	18.20	33.50	30.79	41.10	42.90	39.06	35.29	40.70	37.20	34.16	29.00	24.28		36.22	35.70			43.82	
	Service	21.60	13.10	63.60	65.69	56.50	51.90	57.86	54.00	57.20	58.90	50.72	37.00	40.32		62.20	62.59			45.16	
1985	Agriculture	28.44	62.40	2.40	3.09	1.90	4.50	2.69	8.80	1.70	2.50	12.51	24.89	30.89		0.95	0.69			11.54	28.60
	Industry	42.88	20.79	30.90	28.29	39.00	40.10	38.22	34.90	38.30	31.20	36.10	30.79	25.69		33.44	35.20			45.31	22.10
	Service	28.67	16.79	66.70	68.59	59.00	55.40	59.07	56.00	60.00	64.90	51.38	44.29	43.40		65.59	64.09			43.14	49.29
1990	Agriculture	27.11	60.09	2.10	2.90	1.50	3.50	2.12	7.19	1.44	2.09	8.21	17.89	29.02		0.33		16.81	13.89	8.10	22.79
	Industry	41.34	21.39	27.90	26.39	37.30	39.00	38.05	34.09	31.41	32.29	38.18	35.40	26.49		32.34		48.35	40.00	38.69	22.70
	Service	31.54	18.50	70.10	70.69	61.20	57.60	59.82	58.20	67.14	64.90	53.59	46.70	44.48		67.32		35.03	45.59	53.21	54.50
1995	Agriculture	19.96	52.20	1.61	2.90	1.09	3.20	1.73	5.70	1.57	2.00	5.82	12.40	26.26		0.16	0.20	7.16	15.70	9.85	26.10
	Industry	47.18	23.00	26.31	24.30	32.34	36.00	32.71	33.60	30.28	27.30	38.38	33.30	27.40		33.75	31.00	36.96	34.00	40.00	19.60
	Service	32.86	24.80	72.08	72.90	66.56	60.80	65.56	60.40	68.15	70.20	55.80	54.30	46.34		66.09	68.80	55.88	50.00	50.15	54.30
2000	Agriculture	15.06	50.00	1.19	2.60	1.11	2.60	1.59	5.10	0.90	1.50	4.39	10.60	23.02	59.90	0.10		6.43	14.50	5.60	
	Industry	45.92	22.50	23.44	23.20	30.51	33.50	31.00	31.20	26.80	25.10	38.09	28.10	26.00	16.00	34.83		37.95	28.40	27.73	
	Service	39.02	27.50	75.37	74.30	68.38	63.80	67.42	63.10	72.30	73.10	57.51	61.20	50.98	24.00	65.07		55.62	57.10	66.67	
2005	Agriculture	12.12	44.80	1.21	1.60	0.80	2.40	1.22	4.40	0.63	1.30	3.15	7.90	18.81	55.80	0.06	1.10	4.97	10.20	5.71	20.50
	Industry	47.37	23.80	22.19	20.60	29.30	29.80	28.05	27.90	23.62	22.20	37.50	26.80	28.13	19.00	32.36	21.70	38.08	29.80	29.27	21.40
	Service	40.51	31.40	76.60	77.80	69.90	67.80	70.73	66.40	75.75	76.30	59.36	65.20	53.06	25.20	67.58	77.30	56.96	60.00	65.02	57.90
2006	Agriculture	11.11	42.60	1.04	1.50	0.82	2.30	1.18	4.30	0.59	1.30	2.99	7.70	18.29		0.05	1.30	4.52	10.00	5.48	19.30
	Industry	47.95	25.20	22.24	20.80	30.14	29.70	28.06	28.00	23.62	22.00	36.86	26.30	28.84		31.71	22.10	37.23	29.30	28.75	21.40
	Service	40.94	32.20	76.72	77.70	69.04	68.10	70.76	66.60	75.78	76.40	60.15	66.00	52.87		68.24	76.70	58.25	60.70	65.76	59.10
2007	Agriculture	10.77	40.80	1.13	1.40	0.87	2.20	1.15	3.90	0.63	1.40	2.71	7.60	18.26		0.05	1.10	4.41	9.00	5.56	18.30
	Industry	47.34	26.80	21.99	20.60	30.50	29.90	28.15	27.40	23.01	22.30	37.01	25.50	29.03		29.35	22.50	36.44	29.20	27.81	22.00
	Service	41.89	32.40	76.88	78.00	68.63	67.80	70.70	67.20	76.36	76.00	60.28	66.90	52.71		70.61	76.40	59.15	61.80	66.63	59.50
2008	Agriculture	10.73	39.60	1.22	1.50	0.96	1.80	1.14	3.80	0.69	1.10	2.51	7.20	17.78		0.04	1.20	4.40	8.60	5.91	17.40
	Industry	47.45	27.20	21.13	19.90	30.13	29.30	27.43	26.90	22.63	21.90	36.28	25.00	28.29		27.33	22.50	36.12	28.90	27.90	22.60
	Service	41.82	33.20	77.65	78.60	68.91	68.90	71.43	67.80	76.68	76.60	61.21	67.90	53.93		72.63	76.20	59.48	62.40	66.18	59.70
2009	Agriculture	10.33	38.10	1.12	1.50	0.75	1.70	1.16	3.90	0.58	1.10	2.59	7.00	17.74	52.00	0.04	1.10	4.69	9.70	5.63	17.00
	Industry	46.24	27.80	19.58	17.10	27.82	28.80	25.94	25.90	21.26	19.50	36.68	16.40	27.76	14.00	27.87	21.80	33.64	27.90	26.83	22.10
	Service	43.43	34.10	79.30	80.90	71.43	69.50	72.90	69.00	78.17	78.70	60.73	76.60	54.50	34.00	72.09	77.10	61.67	62.30	67.54	60.70
2010	Agriculture	10.10	36.70	1.19	1.60	0.80	1.60	1.18	3.70	0.68	1.20	2.47	6.60	18.21	51.10	0.04		3.87		5.30	
	Industry	46.67	28.70	19.79	16.70	30.23	28.40	27.46	25.30	21.52	19.10	38.27	17.00	27.16	22.40	27.63		34.70		28.07	
	Service	43.24	34.60	79.02	81.20	68.97	70.00	71.36	69.70	77.80	78.90	59.26	76.40	54.64	26.60	72.33		61.44		66.63	

From 1980 to 2010, the US service GDP share increased from 63.60% to 79% and its employment share increased from about 65.69% to 81.20%.

China's relevant share of service GDP was 21.60% in 1980 to 43.24% in 2010 and employment share increased from 13.10% in 1980 to 34.60% in 2010.

In Germany, service GDP was 56.50% in 1980 and 68.97% in 2010; and service employment increased from 51.90% in 1980 to 70.00% in 2010.

In Japan, the service GDP was 57.86% in 1980 and 71.36% in 2010; the service employment was 54.00% in 1980 and 69.70% in 2010.

In UK, it was 57.20% of service GDP in 1980 and 77.80% in 2010; and it was 58.90% of service employment in 1980 and 78.90% in 2010.

In South Korea, the service GDP was 50.72% in 1980 and 59.26% in 2010; the service employment was 37.00% in 1980 and 76.40% in 2010.

In Singapore, the service GDP was 62.20% in 1980 and 72.09% in 2010; and the service employment was 62.59% in 1980 and 77.10% in 2009.

In India, the service GDP was 40.32% in 1980 and 54.64% in 2010; the service employment was 24.00% in 2005 and 26.60% in 2010.

In Russia, the service GDP was 35.03% in 1990 and 61.67% in 2009; the service employment was 45.59% in 1990 and 62.30%.

In Brazil, the service GDP was 45.16% in 1980 and 67.01% in 2011; the service employment was 49.29% in 1985 and 62.70% in 2011.

Table 4b. Percentage Changes of GDPs and Jobs in Service Industry in Different Time Periods

		China	US	Germany	Japan	UK	Korea	India	Singapore	Russia	Brazil
1980-90											
GDP % Δ	Agriculture	-3.06	-0.80	-0.90	-0.96	-0.66	-6.90	-6.36	-1.23	16.81	-2.91
	Industry	-6.88	-5.60	-3.80	-1.01	-9.29	4.02	2.21	-3.88	48.35	-5.13
	Service	9.94	6.50	4.70	1.96	9.94	2.87	4.16	5.12	35.03	8.05
Job % Δ	Agriculture	-8.60	-0.69	-1.70	-3.20	-0.50	-16.11	0.00		13.89	22.79
	Industry	3.19	-4.40	-3.90	-1.20	-4.91	6.40	0.00		40.00	22.70
	Service	5.40	5.00	5.70	4.20	6.00	9.70	0.00		45.59	54.50
1990-2000											
GDP % Δ	Agriculture	-12.05	-0.91	-0.39	-0.53	-0.54	-3.82	-6.00	-0.23	-10.38	-2.50
	Industry	4.58	-4.46	-6.79	-7.05	-4.61	-0.09	-0.49	2.49	-10.40	-10.96
	Service	7.48	5.27	7.18	7.60	5.16	3.92	6.50	-2.25	20.59	13.46
Job % Δ	Agriculture	-10.09	-0.30	-0.90	-2.09	-0.59	-7.29	59.90		0.61	
	Industry	1.11	-3.19	-5.50	-2.89	-7.19	-7.30	16.00		-11.60	
	Service	9.00	3.61	6.20	4.90	8.20	14.50	24.00		11.51	
2000-2010											
GDP % Δ	Agriculture	-4.97	0.00	-0.31	-0.41	-0.22	-1.92	-4.82	-0.06	-2.57	-0.30
	Industry	0.75	-3.65	-0.28	-3.54	-5.28	0.18	1.16	-7.20	-3.25	0.34
	Service	4.22	3.65	0.59	3.95	5.50	1.74	3.66	7.26	5.82	-0.04
Job % Δ	Agriculture	-13.30	-1.00	-1.00	-1.40	-0.30	-4.00	-8.80			
	Industry	6.20	-6.50	-5.10	-5.90	-6.00	-11.10	6.40			
	Service	7.10	6.90	6.20	6.60	5.80	15.20	2.60			

Table 4b gives the percentage changes of GDP and Jobs between 1980-1990, 1990-2000 and 2000-2010 in three sectors in these selected countries. The data show how jobs and associated GDPs are shifted from other sectors to the service sector.

The above analyses indicate that in around 30 years, in each of these selected counties, its service sector GDP and employment were increased significantly. In developing countries, the percentage increase of the GDP during the time period surpassed its percentage increase of the

service employment. That means the whole economy's productivity was improved when the labor was shifted to the service sector (mostly from the agricultural sector). However, in developed countries, the percentage increase of the service GDP was slowed down clearly while the service employment increased significantly. That indicates that with the maturity of the economy, the shift of labor to the service sector (mostly from manufacturing) lowered a country's productivity.

As many other countries have experienced, such as China and India with their continued fast economic growth, more jobs will be shifted into the service sector. Currently, service employment is about 36% in China and 28% in India. In other well developed countries, this figure is between 70-80%. Even in Brazil, it is above 60%.

However, one needs to consider that both China and India are large populous countries and both are still under the industrialization stage; therefore their agriculture and industry sectors and its developments will remain to be essential to them. As a result, the shifts of labors to the service sector in these two countries in the coming decade will not be so dramatic as happened in western countries. The reasonable expectation will be about 45% of employment in the service sector in China and 38% in India by 2025.

5. China and Others Countries' Service Sector Distribution

Table 5. Service Distributions in Different Countries

	China			US			Germany			Japan			United Kingdom			South Korea			Brazil	
Type of Service/Year	2000	2005	2012	2000	2005	2012	2000	2005	2012	2000	2005	2012	2000	2005	2012	2000	2005	2012	2000	2005
Distributive trade, repairs; transport; accommod., food se	1,646,580	2,882,806	8,481,840	1,651,480	2,002,675	2,385,914	306,593	376,905	438,731	596,584	765,666	872,843	183,097	229,458	267,462	124,637	152,479	197,519	139,866	285,671
	0.43	0.38	0.37	0.23	0.21	0.20	0.24	0.23	0.21	0.27	0.28	0.27	0.28	0.25	0.23	0.30	0.27	0.26	0.20	0.24
Information and communication				545,176	729,989	920,463	79,995	91,951	121,236	163,637	202,767	250,010	61,084	78,772	92,214	32,215	45,956	51,879	44,580	84,399
				0.07	0.08	0.08	0.06	0.06	0.06	0.07	0.07	0.08	0.09	0.09	0.08	0.08	0.08	0.07	0.06	0.07
Financial and insurance activities	408,669	608,683	2,872,268	722,739	952,102	1,074,935	83,205	113,534	119,215	163,520	237,654	204,989	51,065	95,456	119,761	41,759	67,681	87,984	70,590	131,399
	0.11	0.08	0.12	0.10	0.10	0.09	0.06	0.07	0.06	0.07	0.09	0.06	0.08	0.11	0.10	0.10	0.12	0.12	0.10	0.11
Real estate activities	414,906	851,643	2,900,549	983,658	1,339,263	1,705,411	208,075	265,102	365,260	351,486	417,143	540,746	77,835	103,728	167,285	66,192	80,129	93,483	126,160	171,758
	0.11	0.11	0.13	0.14	0.14	0.14	0.16	0.16	0.18	0.16	0.15	0.16	0.12	0.11	0.14	0.16	0.14	0.12	0.18	0.14
Prof., scientific, techn.; admin., support serv. activities				1,074,308	1,361,572	1,768,531	208,509	254,489	333,972				98,549	132,720	175,777	31,516	48,031	71,572	58,213	99,979
	-			0.15	0.14	0.15	0.16	0.16	0.16	-	-	-	0.15	0.15	0.15	0.08	0.08	0.09	0.08	0.08
Public admin.; compulsorys.s.; education; human health				1,939,589	2,601,615	3,460,375	332,385	410,832	553,161	354,661	424,118	519,444	152,871	221,559	281,877	97,644	156,240	227,996	160,787	291,761
				0.27	0.28	0.29	0.25	0.25	0.27	0.16	0.16	0.16	0.23	0.24	0.24	0.24	0.27	0.30	0.23	0.24
Other service activities	1,401,241	3,148,796	8,885,992	358,419	406,548	476,115	84,975	105,032	137,271	577,461	682,609	891,767	33,115	46,058	59,653	16,357	24,001	33,208	103,308	150,408
	0.36	0.42	0.38	0.05	0.04	0.04	0.07	0.06	0.06	0.15	0.25	0.27	0.05	0.05	0.05	0.04	0.04	0.04	0.15	0.12
Total service output	3,871,395	7,491,927	23,140,649	7,275,369	9,393,764	11,791,744	1,303,738	1,617,845	2,068,845	2,207,350	2,729,956	3,279,798	657,616	907,751	1,164,029	410,320	574,517	763,641	703,504	1,215,375

Table 5 lists the GDP in each service sector in these selected countries in 2000, 2005 and 2012. In China, percentage of GDP in "Distributive trade, repairs, transport, accommodation and food service" to its total service GDP was about 40%; the GDP in "Financial and insurance activities" was about 10%; it was about 10% in "Real estate activities" and 36% in "Other services". China does not have relevant statistics in "Information and communication", "Professional scientific, technology, administrative, support service activities", and "Public administration, compulsory social service, education, human health". The lack of GDPs in these three sectors could be due to the differences in classifications. Most likely, these missed services should be included in "Other services". In the US, "Distributive trade, repairs, transport, accommodation and food service" composed about 20% of the total service GDP, and the share of "Public administration, compulsory social service, education, human health" was about 28%. In South Korea, it was about 27% each in "Distributive trade, repairs, transport, accommodation and food service" and "Public administration, compulsory social service, education, human health". Compared with other countries, China's GDP shares in "Information and communication" , "Professional scientific, technology, administrative, support service activities", and "Public administration, compulsory social service, education, human health" were too low.

6. China and Other Countries' Service Exports

Table 6. Service Exports in Different Countries (current US$)

Country/Year	2005	2006	2007	2008	2009	2010	2011	2012	Average
China	$89,149,912,936	$113,851,040,171	$147,510,687,295	$165,990,031,386	$144,184,537,444	$171,489,835,442	$184,763,043,713	$196,301,948,474	$151,655,129,608
United States	$376,499,000,000	$419,502,000,000	$489,540,000,000	$534,030,000,000	$510,623,000,000	$556,987,000,000	$618,317,000,000	$650,564,000,000	$519,507,750,000
Germany	$176,698,839,227	$200,017,774,620	$233,132,519,160	$265,305,258,177	$248,696,914,882	$253,555,322,669	$274,113,958,546	$268,497,877,829	$240,002,308,139
Japan	$101,960,758,643	$109,345,586,363	$121,579,391,059	$141,040,143,456	$120,914,193,843	$131,215,197,969	$137,348,837,476	$134,185,896,985	$124,698,750,724
UK	$207,673,909,145	$236,552,860,391	$288,285,959,581	$287,956,159,705	$239,563,934,300	$256,705,249,989	$294,344,975,105	$292,646,867,763	$262,966,239,497
Korea	$49,744,800,000	$56,842,400,000	$72,994,800,000	$90,634,300,000	$73,580,400,000	$87,282,400,000	$95,257,200,000	$110,853,800,000	$79,648,762,500
Singapore	$46,241,297,527	$58,951,246,119	$73,853,809,907	$89,298,464,572	$81,611,343,566	$100,687,833,469	$116,160,281,533	$119,075,324,919	$85,734,950,201
India	$52,178,951,919	$69,439,848,438	$86,552,459,544	$106,054,239,105	$92,889,486,182	$117,068,311,674	$138,527,915,665	$145,524,596,558	$101,029,476,136
Russia	$28,845,410,000	$35,718,534,250	$43,860,061,000	$57,135,949,289	$45,796,501,700	$49,158,958,000	$58,039,060,000	$62,340,015,000	$47,611,811,155
Brazil	$14,975,857,163	$18,495,034,407	$23,954,250,578	$30,450,526,537	$27,728,347,800	$31,598,901,061	$38,209,117,260	$39,863,624,582	$28,159,457,423

Table 6 provides the total service exports of selected countries during 2005-2012. The US is ranked as #1 in total exports in each of these years, UK #2, and Germany # 3. Although China is #1 exporter in the world, its total service export was only ranked #4, with the average annual service export of $151,655,129,608, which was less than 30% that of the US (average annual service export of $519,507,750,000).

Table 6a. Information, Communication and Technology Service Exports (current US$)

Country/Year	2005	2006	2007	2008	2009	2010	2011	2012	Average
China	22,202,966,679	27,731,770,927	39,111,105,584	47,031,906,970	45,867,524,462	53,641,608,808	60,690,749,652	68,540,636,096	**45,602,283,647**
U.S		78,376,000,000	94,572,000,000	105,185,000,000	107,753,000,000	117,313,000,000	131,949,000,000	143,962,000,000	**111,301,428,571**
Germany	52,439,660,632	59,743,463,711	71,879,241,797	84,268,513,759	80,110,465,631	82,212,168,371	90,128,275,179	90,529,118,463	**76,413,863,443**
Japan	20,238,044,590	23,691,376,650	26,364,949,708	34,388,465,303	36,030,158,696	33,492,592,579	38,588,091,031	27,794,782,335	**30,073,557,612**
U.K	77,986,547,674	87,793,954,826	104,612,379,236	103,817,519,347	90,528,498,865	106,863,708,028	115,778,917,690	114,080,870,961	**100,182,799,578**
Korea	9,922,400,000	11,422,600,000	15,308,000,000	13,818,700,000	12,860,800,000	17,667,700,000	19,471,000,000	23,929,600,000	**15,550,100,000**
Singapore	10,292,072,657	15,568,809,264	17,587,466,451	21,130,064,029	20,397,593,593	23,018,927,888	27,529,467,394	28,532,110,873	**20,507,064,019**
India	35,073,974,538	47,423,914,835	58,985,122,161	73,484,592,821	61,624,368,553	75,037,123,007	85,662,135,486	95,891,807,671	**66,647,879,884**
Russia	6,833,560,000	8,919,170,000	12,366,658,000	16,712,923,441	14,005,980,410	14,965,290,000	17,844,080,000	19,901,820,000	**13,943,685,231**
Brazil	6,435,168,000	7,891,540,000	11,483,100,000	14,971,477,000	14,404,566,000	16,422,484,742	20,198,895,520	22,222,660,386	**14,253,736,456**

Table 6a is the exports in Information, Communications and Technology during 2005-2012. China's annual ICT export was ranked as # 5, less than that of India.

Table 6b. Insurance and Financial Services Exports (% of service exports)

Country/Year	2005	2006	2007	2008	2009	2010	2011	2012	Average
China	**0.78**	**0.62**	**0.77**	**1.02**	**1.36**	**1.78**	**2.09**	**2.66**	**1.38**
U.S	**12.01**	**13.67**	**14.75**	**14.31**	**15.48**	**15.57**	**15.07**	**14.22**	**14.38**
Germany	**4.95**	**6.21**	**7.64**	**6.83**	**7.42**	**7.37**	**7.71**	**7.42**	**6.94**
Japan	**5.80**	**7.07**	**6.21**	**4.53**	**4.68**	**3.72**	**4.18**	**3.16**	**4.92**
U.K	**21.92**	**24.85**	**28.51**	**29.11**	**27.18**	**23.39**	**27.88**	**28.54**	**26.42**
Korea	**3.66**	**4.96**	**6.05**	**4.69**	**3.56**	**3.72**	**4.10**	**3.33**	**4.26**
Singapore	**12.56**	**13.62**	**15.99**	**14.20**	**15.94**	**15.55**	**15.32**	**14.97**	**14.77**
India	**3.99**	**5.00**	**5.64**	**5.52**	**5.52**	**6.50**	**6.38**	**5.23**	**5.47**
Russia	**2.47**	**2.70**	**3.54**	**3.12**	**3.07**	**3.02**	**2.48**	**2.80**	**2.90**
Brazil	**4.28**	**5.74**	**6.82**	**6.78**	**7.01**	**7.88**	**8.29**	**8.09**	**6.86**

Table 6b is the percentage of Insurance and Finance export to the country's total service export during 2005-2012 for these selected countries. China has the lowest percentage with only

1.38% among all of these selected countries. UK is ranked as #1 with 26.42%, Singapore #2 with 14.77%, and US # 3 with 14.38%. Even other BRIC countries--Russia with 2.90%, India with 5.47% and Brazil with 6.86 % -- were much higher than China's percentage.

Through the above analyses, it's clear that China needs to improve its service industry's competitiveness so that it can export more services. In particular, China needs to strengthen its services in sectors of Information and Communications and Insurance and Finance. More Chinese companies are doing business in other countries, and China's FDI outwards have been increasing over the years. There should be more demand for relevant services in overseas from Chinese companies, especially in the areas of insurance, finance and professional services such as legal and accounting.

IV. What China Can Learn from Other Countries and What it Should Do

China is at the crucial stage of its economic development. It needs to identify new sources and engines to sustain its stable growth since previous approaches using mass exports and government investments have reached their limit. Stimulating domestic consumption and better developing its service industry have become China's new directions.

As outlined in my paper (Chen, 2014)), currently manufacturing exports comprise 93% of China's total exports. Worldwide, that number is about 70% of total exports. Given China's development stage and employment needs, China still needs to continuously improve its manufacturing industry and export enough such products. In particular, China needs to become the strongest manufacturing country to sustain its economic growth. Interestingly, Prime Minister Li Keqiang used the same words to emphasize the importance and improvements of China's manufacturing industry in its annual report to China's Congress in March this year. He further outlined China's Manufacturing Industry 2025 similar to Germany's Manufacturing 4.0.

Besides stabilizing and improving its manufacturing industry, China certainly needs to expand and strengthen its service sector. As discussed in the previous sections, increasing importance and particular growing shares of employment and GDPs in services have been the worldwide trend that many developed countries have experienced. The issues germane to China are: how fast the service sector development can be and what will be potential problems, and how to solve them if service development is too fast.

Currently, China's service GDP is about 40% of its total GDP and employment is about 30% of total employment. In the past 12 years, China's annual GDP share in service rose by 0.47% and job share by 0.61%. By 2025, China should expect about 50% GDP and 45% jobs in its service. By 2035, service GDP should be around 55% and employment around 52%.

China's Expected GDP and Jobs in 2025 and 2035

	2012		2025		2035	
	GDP %	Job %	GDP %	Job %	GDP %	Job %
Agriculture	10.08	33.6	8	25	5	20
Industry	45.27	30.3	42	30	40	28
Service	44.65	36.1	50	45	55	52

There will be some serious outcomes and problems if China grows its service sector too fast, i.e. its GDP and employment shares in services are too high, compared with the above proposed ranges.

1. *Weightless.* As some western countries have experienced, manufacturing outsourcing led to their economies to be weightless. The US government has called for manufacturers to come back to the US. If its GDP share in manufacturing becomes too low (less than 40%, or even lower than 35%), China' economy will face the weightless problem. Many better paid jobs in manufacturing will be lost and outsourced to other developing countries.
2. *Cost Crises.* As Baumol (1967) stated that shifting jobs from manufacturing to service may cause the cost crisis. The productivity in service is not rising but the average salary must be raised because of job market competition and the productivity's improvement in manufacturing, so the salary rises there. If such a crisis occurs, service firms will suffer and that will further negatively affect the whole economy.
3. *Overall Productivity.* When jobs shift from agriculture to manufacturing, the whole economy's productivity improves significantly. When too many jobs shift from manufacturing to service, the average productivity will be lower.
4. *Low Salary.* Usually, average salary in service is lower than in manufacturing. When many people move from manufacturing to service, their salary and income will be lower. That has been the case in many developed countries.
5. *Competitiveness.* The competitiveness is essential to an economy's strength. Globalization makes a country's competitiveness more important. Overdevelopment and shrinking of manufacturing sector will lead China to lose its competitiveness in the world. What are China's competitive advantages? More in manufacturing or in service?
6. *Exports.* Given China's large population and economic development stage, having enough exports will still be crucial to China's economy in coming decades, although contributions of the exports to China's GDP should be gradually lower. For more than two decades, China relied on increasing exports to support its rapid economic growths. In some years, the ratio of export to its GDP was above 40%. In the future, China needs to rely more and more on its domestic consumption. But China still needs to maintain its fair share of the world's exports. If China cannot develop and implement an appropriate economic structure strategy and improve and strengthen both manufacturing and service sectors, it then will lose more world market share.
7. *Mid-income trap.* When a country uses its competitive advantages to advance its economy and reaches a certain income level, it may get stuck at that level

because these previous competitive advantages have become obsolete and new ones could not be created. That is called the 'mid-income trap' as some newly industrialized countries like South Africa and Brazil have experienced. Their per capita GDP has remained around $10,000 for more than a decade. These countries have suffered from low investment, slow growth in manufacturing, limited industrial diversification and poor labor market conditions. China's GDP per capita at present is around $7,000 and will reach $10,000 in about 6 years. To avoid such a mid-income trap will be a big challenge to China. Too rapid an expansion in service and shrinking in manufacturing will lead to such a trap, not solving the potential problem.

What should China do in order to maintain balanced, stable and sustainable economic development? As Chen (2014) has emphasized, China must continuously strengthen and improve its manufacturing industry and make it the strongest sector. It needs to diversify its manufacturing industry and particularly develop sectors that consume less natural resources and emit less pollution. Also, it should develop more manufacturing products that can be used to directly substitute current mass imports and the ones representing future trends.

Expanding and developing the service industry will be an irreversible trend. Eventually, the GDP in service will surpass 50% of its total GDP and jobs in service will be above 50% of the total employment. But the issues facing China will be how soon that will be, what types of services will be needed, and what types of jobs will be necessary to sustain this growth.

As discussed in the previous sections, growing of the service sector and especially shrinking of manufacturing should be gradual, not dramatic. Social services and insurance/financial service should be given the highest priority. China now has more than 200 million seniors. Providing more and better services to these seniors should be the government's and the society's responsibilities and that also creates large business opportunities. Insurance and financial service is quite underdeveloped in China and lags greatly when compared with other developed countries. Professional services, including legal, accounting, medical and psychological services, has huge potential for growth. Also, China needs to expand and improve its manufacturing services.

Another area is service export. As pointed out in the previous discussions, China's service export is far behind many other countries. Currently only about 7% of its total export is in service. Raising this ratio to 12-15% of its total export in the future should be reasonable and achievable.

V. Conclusions

This paper provides a comprehensive study and international comparisons of the service industry throughout the world. The data of the economic structure changes over the years for different countries, particularly the changes of the percentages of service GDPs and service employments in these countries show that the path of service industry development is based on each country's economic development level as well as its unique conditions. But generally speaking, the service industry becomes increasingly important when an economy advances.

In order to avoid the potential mid-income trap, a country needs to continuously improve its manufacturing industry and particularly its competitiveness and diversity. It needs to develop and expand the service sectors which are not only highly needed but also productive and skill-

intensive. Otherwise, the low average productivity and low average wage will slow the aggregate demand and with it the whole economy.

China is facing huge challenges and its economy is at a turning point. In order to maintain stable and sustainable economic growth, China needs to adjust its economic structure and better develop the service industry. It especially needs to expand and develop its professional, social and manufacturing services to better support its development.

China needs to adjust its economic development strategy and rely more on domestic consumption. It needs to focus more on innovations and institutional reforms to sustain economic growth. Sustaining and improving international trade and particularly exports will still be essential to China's economy and its future developments. Creating competitive advantages of its service industry and exporting more services will be crucial to China's future success.

Acknowledgement and Data Sources

The author wants to thank University of the West for providing financial assistance to this research project (through the UWest Faculty Research Grant) and also wants to thank UWest students Adam Yuan and Kathleen Wu for their fine research assistance. This paper was presented at the 25th International Conference on the Pacific Rim Management, July 9-11th, 2015, Los Angeles. The author wants to thank the participants of the conference for their valuable comments.

Unless otherwise mentioned, all data sources used in this paper are from the World Bank at www.worldbank.org.

References

Baumol, William J (1967), " Macroeconomics of unbalanced growth: the anatomy of urban crisis." ***Am Econ Rev*** 57: pp. 415-426.

Baumol, William. J.; Blackman, Sue. Ann. Batey.; Wolf, Edward. N. (1989). *Productivity and American Leadership: The long view.* Cambridge, Massachusetts, MIT Press.

Bell, Daniel. (1973). ***The Coming of postindustrial society: A venture in social forecasting***. New York, Basic Books. (Consulted in the 1976 edition, with a new introduction by the author.)

Castells, Manuel and Aoyama, Yoku. (1993), "*Paths towards the informational society: A comparative analysis of the transformation of employment structure in the G-7 countries, 1920-2005."* **BRIE** Working Paper No. 61. Berkeley Roundtable on the International Economy, University of California, Berkeley.

Chen, Yueyun (2014), "China's Path to Sustainable, Stable and Rapid Economic Development: From the Largest to the Strongest Manufacturing Country," ***Academic Perspective***, Volume 10, pp 84-106 .

Clark, Colin. (1957), ***The Conditions of Economic Progress***, London: Macmillan.

Cohen, Stephen; Zysman, John. (1987). *Manufacturing matters. The myth of the postindustrial economy.* New York, Basic Books.

Conte A. and M. Vivarelli (2007), "Globalization and Employment: Imported Skill Biased Technological Change in Developing Countries," Jena Economic Research Papers 2007-009, Jena.

Denison, Edward. (1985). ***Trends in American economic growth, 1929-1982***. Washington. DC, Brookings Institution.

Dietrich Andreas (2009), “Does growth cause structural change, or is it the other way around? A dynamic panel data analysis for seven OECD countries.” **Jena Economic Research Papers in Economics** 2009-034.
Fisher, Allan GB (1939), “Production, primary, secondary and tertiary.” **Econ Rec** 15: pp. 22-38.
Hansen, Gary D. and Prescott, Edward C. (2002): ‘Malthus to show’, ***American Economic Review***, Vol. 92 (4), pp. 1205-1217.
Krüger, Jens J (2008), “Productivity and structural change: a review of the literature.” **J Econ Surv** 22: pp. 330-363.
Kuznets, Simon (1966), ***Modern Economic Growth. Rate Structure and Spread***, New Haven and London: Yale University Press.
Lewis, W. Arthur (1972), “ Reflections on Unlimited Labor,” in L. E. DiMarco (ed.) International Economics and Development (Essay in Honor of Raoul Prebisch), New York: Academic Press, pp. 75-96.
Lucas, Robert E (1988), “On the mechanics of economic development.” **J Monetary Econ** 22: pp. 3-42.
McKinsey & Company, Global Institute (2012) **“**Manufacturing the Future: The Next Era of Global Growth and Innovation.**”**
Monk, Peter (1989), ***Technological change in the information economy*****.** London, Pinter.
Nelson, Richard (1981), “‘Research of productivity growth and productivity differences: Dead ends and new departures,” ***Journal of Economic Literature***, 19/99:1029-1064.
Oulton, Nicholas (2001), “Must the Unbalanced Growth Rate Decline? Baumol’s Unbalanced Growth Rate Revisited,” ***Oxford Economic Papers***, Vol. 53 (4), pp. 605-627.
Pelka, GW (2005), Wachstum und Strukturwandel. Metropolis-Verlag, Marburg.
Porat, M. (1977). ***The information economy***, Washington, DC, US Department of Commerce, Office of Telecommunications.
Potts, Jason and Tom Mandeville (2007), “ Toward an Evolutionary Theory of Innovation and Growth in the Service Economy,” **Prometheus**, vol. 25, issue 2, pages 147-159.
Romer, Paul M (1986), “ Increasing returns and long-run growth.” **J Political Econ** 94: pp. 1002-1037.
Romer, Paul M (1990), “Endogenous technological change.” **J Political Econ** 98, no. 5. Part 2: the problem of development: a conference of the institute for the study of free enterprise systems, pp S71–S103.
Rostow, Walt Whitman (1960), "The Five Stages of Growth-A Summary". The Stages of Economic Growth: A Non-Communist Manifesto. Cambridge: Cambridge University Press. pp. 4–16.
Sautter, Christian. (1976), “Lefficacite et la rentabilite de leconomie francaise de 1954 a 1974,” ***Economie et Statistique*** (Paris), p.68.
Silva, Eater G, Teixeira, Aurora AC (2008), “Surveying structural change: seminal contributions and a bibliometric account.” **Structure Chang Econ Dyn** 19: pp. 273-300.
Smith Adam (1776), An inquiry into the nature and causes of the wealth of nations. In: Campbell RH, Skinner AS, Todd WB (eds) Reprinted in 2 vols. Clarendon Press, Oxford.
Solow, Robert M (1956), “A contribution to the theory of economic growth.” **Q J Econ** 70: pp. 65-94.
Solow, Robert M (1957), “Technical change and the aggregate production function.” **Rev Econ Stat** 39: pp. 312-320.
Stiglitz, Joseph and Linda J. Bilmes (2012), “The Book of Jobs,” Vanity Fair.
Touraine, Alain. (1969). ***La societe post-industrielle***. Paris, Denoel.

Turgot ARJ (1766), Reflections on the formation and the distribution of wealth. In: Meek RL (eds): ***Turgot on Progress, sociology and economics.*** Cambridge University Press, Cambridge, pp 119–134

电影中的东亚有组织犯罪

夏昀

Valparaiso University, IN, USA

引言

古往今来，有组织犯罪集团在几乎所有社会和文化背景中都有其存在空间。 政府，执法机关，研究人员和普通民众对于有组织犯罪的理解侧重点也不尽相同。根据1967年美国针对有组织犯罪特别行动组的定义，有组织犯罪（organized crime）指的是“系统地向大量普通民众提供赌博，高利贷，毒品等非法物品和服务的暴力团伙。”[i] 此定义强调的是有组织犯罪集团赖以牟利的非法交易内容及其在社会中长期存在的合理性。此类犯罪分子满足的是社会上很多难以从正常途径满足的需求。除了上述定义列出的物品和服务以外，还有保护，逼债，恐吓，暗杀及性服务等等。民众印象中的“黑帮” 或者“黑道”，也不尽是有组织犯罪。例如，1949年国民政府迁台以后，在孤岛上长大的外省子弟为了寻求归属感及免受当地人欺负，盟誓建立帮会组织。后来大名鼎鼎，势力遍及台湾政商及娱乐业的竹联帮及四海帮即成立于此背景之下。外省帮会成立之前，台湾本省人也有自己的组织， 即“角头”。角头数量众多，结构松散，各自占有一片区域，以保护费为主要收入来源，兼营规模有限的酒店和赌场。在国民政府司法院的考量中，大部分典型的“角头” 都不在有组织犯罪之列。[ii]

有组织犯罪是大部分普通民众没有机会也不愿碰触的领域，是以中文有“黑社会”“黑道”和“黑社会”之称，而日文中也成为“极道”。相对于“黑”，或“极”的，是大部分人工作生活的合法空间和主流社会。正是因为缺乏其他了解渠道，反映有组织犯罪的电影在美国，欧洲及东亚一向是商业电影中最经典和最卖座的类别之一。银幕上风度翩翩，拳脚利落，又往往重情重义、盗亦有道的黑帮形象影响了数以亿计的观众，尤其成为年轻人模仿追捧的对象。甚至连真正的黑道中人也从电影中获得灵感。学者Federico Varese指出，意大利黑手党乐于见到电影中对于他们暴戾狠辣的形象的塑造。在进行不法交易或威胁受害者时，他们甚至模仿电影中黑手党头领的举止和台词。对方心惊胆战之时，他们反而不需要真正诉诸暴力。[iii]

电影作为一种大众媒介，因其娱乐性和凝练逼真的叙述方式，是对于有组织犯罪这一大众好奇却陌生的世界的最佳展现渠道。东亚有组织犯罪根据其产生地区和文化背景不同而各具特色，其在银幕上的展现也是如此。同时，东亚各国的黑帮电影也表现出一些在区域政治、经济和影业大环境影响下的共同特征。本文将以日本，香港和台湾黑帮电影为主，简单探讨这一银幕类型的主题，特点和变迁。虽然中国大陆的观众对于这些电影中的很多都耳熟能详，但大陆地区本身涉及有组织犯罪的电影极其有限。近年来，韩国在这一类型中佳作频出，但由于篇幅所限，也不在讨论之列。

一． 黑道电影的黄金时期

日本黑道电影 （yakuza films）在二十世纪20年代即出现，真正崛起却是在二战后，于60-70年代达到黄金时期。到了80年代，电影本身面临电视这一新媒体的冲击而显颓势，90年代又凭新一代导演大胆创新之作重新占据市场。香港电影在80年代末至90年代达到数量和质量的高峰， 这一时期的黑道电影，如刘伟强导演的古惑仔系列，也跻身最卖座的影片。黄金时期的黑帮电影在日本和香港各有其鲜明的主题和情节重点，并且

都在黑道人物身上寄托了当时的文化和社会背景中普通人不能达成的目标和理想。日本经典黑道电影的主角往往沉默而秉持原则，却身不由己地陷入某个事件或危机中，而不得不在“人情”（*ninjo*）和“义理”（*giri*）中取舍。[iv] 深作欣二在名作《无仁义的战争》系列中塑造的广能昌三，尤其在第二部《广岛死斗》中，就体现出这个特点。[v] 这个情节套路是如此充满戏剧张力而极具日本特色，以至于 1975 年美国导演 Sydney Pollack 的作品《极道》（《The Yakuza》）中，白人主角在他的日本黑道友人的启示下，也在“人情”和“义理”的挣扎中做出了符合日本道德文化判断的选择。

香港黄金时期的黑道电影也有一个突出主题，即忠义。华人帮派的起源，组成和变迁与中国近现代历史中的转折性事件密不可分。很多学者认为清朝的白莲教和天地会等秘密会社即是现代黑社会的鼻祖。[vi] 鸦片战争以降，华人开始四散海外，在各地唐人街初步形成时期，秘密会社也在异国土壤发展，并对清末起义和革命起到推动和支持的作用。可以说，从高举反清复明旗帜开始，秘密会社（或当时政府眼中的有组织犯罪）即建立起有民族责任感和爱国的形象。20 世纪中叶以来，香港成为华人有组织犯罪集团最为集中也最为高调的地区。90 年代的香港黑帮电影更是制造出一批满足商业社会审美要求且具备中国传统男性道德观的首领形象。“忠”、“义”是这一时期电影主角不可缺少的性格特点。关公作为中国历史文化中忠义的代表人物，其塑像几乎出现在所有黑帮电影中，道上兄弟拈香拜关公的场景也屡见不鲜。《古惑仔》系列塑造了对流行文化影响深远的少年黑帮头领 -- 陈浩南。其信守承诺，勇于担当，对于跟随他的兄弟和扶持他的黑道大哥的义气，都使这个游走于社会边缘和法律之外的人物具备了商业电影中正面人物的吸引力。《古惑仔》系列以青少年帮派分子讲述江湖险恶的视角及对义气的强调对华人电影影响深远。2010 年的台湾电影《艋舺》即借用了同样的视角，但对义气这个主题做了更为现实残酷的解读。

当日本和香港电影分别渡过黄金期之后， 随着 20 世纪末全球化步伐的加快和电影工业的多元化，黑帮电影开始渐渐远离理想化，在人物塑造上更加复杂写实，在电影风格上更加黑暗，夸张和碎片化。首先，日本和香港都各自涌现出一批深受后现代思潮影响，并受到本土以外的电影人和观众关注的导演。日本以北野武和三池崇史为例，都以毫不克制的暴力画面，对边缘人物的关注和冷静而有些许夸张的叙事著称。香港自 2000 年以来，除了在 90 年代引领黑帮片风潮的刘伟强继续佳作不断外，尔冬升和杜琪峰尤以质量和数量取胜。下述分析将着眼于黄金时期以来日本和香港黑帮电影的变迁，并以这几位导演的作品为主，穿插对其他相关电影的讨论。

二． 从本土到多元化

黄金时期的日本和香港电影都体现出强烈的本土特点。日本和香港都是多元社会，但日本在平成时代以前，正如香港在 97 之前，电影多关注本土的事件，思潮及人物，并体现出不同程度的民族主义和本土主义倾向。黑帮电影也不例外。深作欣二以手拿摄影机拍摄《不仁义的战争》，以纪录片的真实感记录战后民生凋敝，在美国军事管制下的日本社会。[vii] 在美国大兵肆意污辱日本妇女的时候，只有后来成为黑道首领的人物强行干涉，维护了战败的日本所丧失的男性气质和尊严。这个片段当中的对话和动作，简介有力地表现出日本黑道后来在政治上保守，偏向右翼，民族情绪强烈等特点。

90 年代末期，全球化浪潮席卷东亚，并随着几个重要转折事件的发生，尤为深刻地影响了商品和人口在地区之间的流通。日本长期以来有着排外的名声以及苛刻的移民政

策，但随着经济的衰退和人口老龄化，在吸收外来移民方面越来越宽松。随着 80 年代末期台湾政治民主化和中国大陆改革开放政策，大量移民和非法移民由这两地涌入日本。日本黑道也在系统地从国外吸收廉价劳工并组织偷渡方面扮演相当重要的角色。 而香港向来都是亚洲经济欠发达国家底层群体的寓居，避难和发财之地。学者 Gordon Mathews 称位于九龙的重庆大厦为“低端全球化的枢纽”（a hub of ‘low-end’ globalization）。[viii] 在回归中华人民共和国以后，来自内地的短期居住者和移民更是与日俱增。外来人口，尤其是移民中的不安定成分对于本地社会，包括黑社会的冲击， 都在 90 年代末期以来的日本和香港电影中有所体现。

一个明显的标志就是，在黑帮电影中首次出现了外来者的主角，例如三池崇史 1995 年的作品《新宿黑社会》和尔冬升 2004 年的作品《旺角黑夜》。《新宿黑社会》的副标题为“中国黑帮的战争”，是三池崇史无国籍黑帮电影系列的第一步。主人公之一王志明来自台湾，是龙爪帮的首领。而片中几乎所有人物，包括主角桐谷龙仁，都由于复杂的血统或背景而饱为认同危机所扰。龙仁的父亲是二战后残留中国的日本孤儿，而龙仁作为中日混血儿，一直在警署处于被边缘化的地位，并常做出有悖法律和职业道德的行为。故事的主要情节在东京新宿地区展开，又延伸到台北和台湾乡下，片中多种语言，尤其是日语和台语的大量运用，体现出国界和文化界限的流动性。[ix] 新宿作为大量移民和偷渡客的集中区，被多位导演选址，以表现和批判多文化性（multiculturalism）。《新宿黑社会》中，王志明领导的龙爪帮与来自中国大陆的帮派和本地黑社会及警察之间角逐。 黑帮对于新宿弹丸之地的争夺，在本质上体现的是有共通性而又矛盾深远的东亚各文化之间的碰撞和竞争。现实世界中，日本黑帮已经被崛起的“三国帮派”（“三国”是日本二战期间对处于其殖民统治下的中国大陆，韩国和台湾的合称）挤出新宿。[x] 多文化性对于当地社会和黑帮本身的冲击被精确地投射到电影当中。

同样的现象出现在香港黑帮电影中，只是由于香港与中国大陆在地理上的临近和政治上的从属关系，使来自内地的群体在香港电影中的多元化世界中占据越来越重要的位置。《旺角黑夜》的男女主角分别是来自内地的杀手来福和卖春女丹丹。来福被先来香港定居的同乡老六安排接下杀人生意而卷入香港两个帮派之间的争斗。这一情节设置突出体现了现今暴力集团跨境交易的特征及和集团贯穿上下，向外延伸的等级制度。杜琪峰的《黑社会》系列，则反映出香港黑帮在中国大陆开疆拓土，往返两地，涉足房地产和毒品交易等现实中的问题即两地执法机关面临的挑战。 另外，内地的执法人员也在香港黑帮片中扮演愈加重要的角色， 如《毒战》中为了抓捕香港毒枭全军覆没的内地警察，《黑社会：以和为贵》中的公安厅副厅长等。[xi] 这些电影同时反映出，影响香港黑帮派系斗争和权利更迭的不再只是本地的人物和事件；内地的政治政策，权力人物的插手，甚至作为卒子的底层人物，都可能改变整个格局。

三．从趋义到趋利，从义气到背叛

黄金时期后的黑帮电影的第二个特点，是摒弃了之前经典电影中对主角的理想化设定，更加直白地揭露有组织犯罪内部为了争夺权力、获取利益互相残杀的黑暗面和肮脏手段。日本和香港的黑帮本来各有一套行为准则，而经典的黑帮电影也塑造出一批不惜为原则，组织或兄弟牺牲的人物。日本的黑道在幕府末期兴起，最早的成员中有很多是失去生活来源的流浪武士。 因此日本的黑帮分子仿效武士，发展出一套以正义和责任为核心的处世准则（*jingji*）,包括不骚扰平民， 甚至在灾害时期为人民提供援助和救济等等。[xii]

而香港的黑帮，如三合会等，也有一套严格的入会仪式，要求新会员烧香盟誓，遵守洪门准则。准则的核心即待帮众如手足，遵帮派为家庭，绝不出卖兄弟和组织等等。[xiii] 经典的日本和香港黑帮电影中塑造了许多有正义感且重情重义的孤胆英雄，也产生了一批给角色注入生命力，又借这些角色深入人心的电影明星。 石井辉南 1965 年的作品《网走番外地》，是二战后第一部大卖的黑帮片，饰演主角的高仓健，也凭此片大红大紫。 高仓健塑造的冷峻隐忍的黑帮形象可说是升华了这一时期人们对于黑帮人物的想象。同样的，吴宇森 80 年代末开香港黑帮片先河的《英雄本色》系列中，周润发塑造的潇洒仗义的小马哥影响了中国两岸三地一代人对于黑道人物的印象。90 年代古惑仔系列的主角们虽然形象更加多元化，少了一些悲剧色彩，但第一主角陈浩南仍体现出这类经典银幕形象最重要的特质。

现代化进程在不同时期影响了东亚各国和地区，有组织犯罪的获利途径也在过去的几十年中有了极大的改变。这些趋势不但在电影中有所体现，黑帮电影对黑帮的权力争斗和逐利本质也有了更加具体现实的描述。从日本，台湾和香港来看，各地的有组织犯罪都经历了人员和组织上的扩张，逐步实现了企业化，并不同程度地与执法机关和政治人物勾结。[xiv] 从最初的受保护费，经营赌场，敲诈勒索，到后来从跨境毒品交易中谋取巨大利润，再到投资娱乐、房地产及其他高回报率产业，在每一个决定和行动都涉及巨大利润得失的情况下，黑帮人物因私利背叛组织和兄弟情义已是常态。

在杨德昌以二十世纪 60 年代的台湾为背景的《牯岭街少年杀人事件》中， 就描述了帮派分子所遵循的传统原则是如何在商业社会的冲击下土邦瓦解的。 片中“小公园帮”的头目 Honey 虽不是主角，却是最理想化，最有骑士外表与精神的少年帮派分子。在他因涉入命案而藏匿台南的一段时间，台北的帮派从组织和价值观上都发生了变化。是以 Honey 评论道，“怎么去了一趟南部，你们这些台北的都开始流行搞钱了。”[xv] 这样一个理想人物，却在单枪匹马地去找对手帮派摊牌时，被对方头目冷不防推到迎面而来的卡车轮下惨死。杨德昌对于当时台湾的帮派乃至整个社会的观察，已经在 Honey 这个恪守传统价值而死的角色上淋漓尽致地体现出来。另一位台湾导演钮承泽则在《艋舺》中反思了 80 年代本省角头所面临的危机，并借此表现了整个台湾当时迷茫的身份认同。艋舺是台北从晚清到日据时代都最为繁华的商业区，也是传统上由本省帮派控制的地方。 数十年以来，角头大哥们安于自己的地盘和生意，并保留了日据时代的观念和生活习惯， 连在帮派械斗中都只用冷兵器，对“西方人带来的，”“下等人才用的”手枪弃之如敝履。正如故事的发展所揭示的那样， 结构松散，小本经营，不思进取的角头们终将被组织严密，高度纪律化和企业化的外省帮派所取代，被迫让出自己的生存空间和文化空间。而在电影中给角头大哥致命一击的，却是看到了这一趋势，顺时而动的本帮中人。

虽然所处地区和文化不同，日本和香港的黑帮也经历了相似的冲击和更新换代。我们于是在近些年的黑帮电影中看到对全球化背景下能够在犯罪组织中维持地位和利润所需要的魄力，冷酷和阴暗手段。学者，新闻人员和各国政府对于黑帮组织的调查，也有助于外界了解这个相对封闭，自成一体的生态系统，使电影的情节设置更加真实。[xvi] 以日本而言，黑帮组织形态和价值观的转变被崇尚后现代解构和暴力美学的导演刻画地分外彻底和令人绝望。 90 年代中期，三池崇史即在黑社会系列中展示黑帮的勾心斗角及肮脏生意。《新宿黑社会》中，台湾的龙爪帮从人体器官的跨境交易中盈利；日本黑帮不但收买警察，其头目猥琐的私生活也与早期电影中的英雄形象大相径庭。 北野武 2010 年的作品 *Outrage*（中译《极恶非道》）从名字上就体现出电影对于传统上黑道所宣扬的仁义和义

气的全面颠覆。在日本版的海报上，更是打出“全员恶人”的宣传语。的确，片中的主要帮会山王会，从会长到下属，都无不金钱至上，上下同级之间全无信任或情谊可言。山王会对于势力弱小的帮派的阴谋吞并，更与商战中大鱼吃小鱼的模式如出一辙。

黄金时期后的香港电影中，背叛也代替了兄弟情义成为情节的重点甚至电影主题。在1988年的《旺角卡门》和2004年的《江湖》中，两位同样的演员（刘德华、张学友）饰演两对经历患难的结义兄弟。前者中，年轻的大哥和小弟虽然常有龃龉，却可以在关键时候为了对方舍弃性命；而在《江湖》中，两人饰演的是已经功成名就的大哥和二当家，却互相猜疑，甚至要买凶除掉对方。杜琪峰的《黑社会》系列更是展示了同门之间为了争夺话事人之权如何不择手段，将曾经盟誓结义的兄弟赶尽杀绝。 2002年香港的救市之作《无间道》及其续集，主要情节和矛盾也沿着不同关系和不同情境中的背叛展开。 香港导演尔冬升描写新宿复杂而等级分明的黑暗世界的作品《新宿事件》，更是将各国帮派之间，以及各帮派内部的种种冲突和背叛进行了全景展示。 日本黑道三和会（此三和会为虚构名字，不同于香港的三合会）的内部矛盾上升到互相暗杀的地步，而三和会的两派又分别利用想在新宿站稳脚跟的中国大陆和台湾的黑社会。在贩毒的巨大利润吸引下，中国大陆帮派内部四分五裂，不再听从带领他们打下江山的大哥的调度，最终在日本黑帮和台南派的夹击下全军覆没。

三． 黑道与警察： 从针锋相对到互为表里

早期的黑帮电影中也有描写黑道与政客和警察中腐败分子勾结的情节。 但在黑道电影多元化以来，对于黑白两道关系的刻画更为深刻复杂，甚至某种程度上，黑道是白道种种黑暗交易和潜规则的隐喻。《新宿黑社会》中，警察所持的种族主义偏见，对人命的漠视和行为的肮脏可与黑道比肩。《无间道》通过黑道与警方互相派出卧底的情节，将站在法律两边的势力对照描写，甚至点出了两者的某种共生关系。 卧底情节也出现在尔冬升2007年作品《门徒》之中。 这些类似的情节诚然共同反映了香港警方在现实中打压黑社会时所用的策略及衍生的问题。[xvii] 另一方面，卧底们普遍陷入的身份认同危机以及被黑社会同化的外表，举止和行为，也传达给观众一种对新的认知方式。 这个世界和其中的人物并不是非黑即白，黑白之间不仅有广大的灰色地带，甚至可以互相影响，互相转化。《新宿事件》中虽无卧底，日本警察与来自中国内地的偷渡客铁头之间惺惺相惜，互相扶持的情节也体现出这种看似模糊的是非观中的人性之处。

结语

本文对黄金时期以来日本和香港反映有组织犯罪的电影特点及其变迁做了简要分析。黑帮电影从理想化，本土化的类型电影发展到如今，仍在商业电影之列，但表现出更加写实，复杂，和多元化的特点。而这些新的特点与东亚各地区电影人之间的合作和互相借鉴，票房的互通共享，及观影经验丰富而日渐挑剔的观众紧密相关。经过几十年的探索和成败，今日的黑帮电影难有石破天惊的里程碑式作品，但普遍水平并不较黄金时期有所下降，反而更加直接真实地反映有组织犯罪的世界，并有更显著的政治和社会批判作用。

参考書目

Adelstein, Jake. *Tokyo Vice: An American Reporter on the Police Beat in Japan*. Vintage Crime, 2010.

Chin, Ko-Lin. *Heijin: Organized Crime, Business, and Politics in Taiwan*. M.E.

Sharpe, 2003.
D. Chris. *Gun and Sword: An Encyclopedia of Japanese Gangster Films*. Poison Fang
Books, 2013.
Dillon, Mike. “The Immigrant and the Yakuza: Gangscapes in Miike Takashi’s DOA.” *Studies in the Humanities* 39/40 (2012-2013), 193-232.
Hill, Peter B.E. *The Japanese Mafia*: *Yakuza, Law and the State.* Oxford University
Press, 2003.
Sharon Ingrid Kwok and T. Wing Lo, “Anti-triad legislation in Hong Kong: issues,
problems and development,” *Trends in Organized Crime* 16 (2013): 74-94.
Lintner, Bertil. “Chinese Organised Crime,” *Global Crime* 6:1 (2004), 84-96.
Mathews, Gorden. *Ghetto at the Center of the World: Chungking Mansions, Hong Kong*. University of Chicago Press, 2011.
Rankin, Andrew. “Recent Trends in Organized Crime in Japan: Yakuza vs The Police and Foreign Crime Gangs.” *The Asia-Pacific Journal* 7.1-2 (2012).
Schilling, Mark. *The Yakuza Movie Book: A Guide to Japanese Gangster Films*. Albany: Stone Bridge Press, 2003.
Varese, Federico. “The Secret History of Japanese Cinema: The Yakuza Movies.”
Global Crime 7:1 (2006), 105-124.

尾注

[i] Hill, *The Japanese Mafia*, 8.
[ii] Chin, *Heijin*, 11.
[iii] Varese, “The Secret History of Japanese Cinema,” 108.
[iv] Chris D. *Gun and Sword*, 23.
[v] 《不仁义的战争：广岛死斗》(仁義なき戦い 広島死闘篇), 深作欣二导演，1973。

[vi] Lintner, “Chinese Organised Crime,” 87.
[vii] 《不仁义的战争》(仁義なき戦い), 深作欣二导演，1973。
[viii] Mathews, *Ghetto at the Center of the World*, 1.
[ix] Dillon, “The Immigrant and the Yakuza: Gangscapes in Miike Takashi’s DOA.”

[x] Rankin, “21-Century Yakuza: Recent Trends in Organised Crime, Part I,” 13.
[xi] 《毒战》，杜琪峰导演，2013 年。 《黑社会以和为贵》，杜琪峰导演，2006 年。
[xii] Schilling, *Yakuza Movie Book*, 20.
[xiii] Linter, Lintner, “Chinese Organised Crime,” 87.
[xiv] Chin, *Heijin*; Hill, *The Japanese Mafia*. 关于黑帮收入来源的变化，见 Rankin, “21-Century Yakuza: Recent Trends in Organised Crime”,
[xv] 《牯岭街少年杀人事件》，杨德昌导演，1991。
[xvi] 如美国记者 Jake Adelstein 基于为《读卖新闻》做记者期间掌握的资料而写成的纪实文学即包含大量关于日本极道的信息。Adelstein, *Tokyo Vice.*
[xvii] Kwok and Lo, “Anti-triad legislation in Hong Kong: issues, problems and development,” 88.