

SI 741 - Systems, Networks, and Webs  
Short Paper - Comparative Theories of Infrastructure  
In matters of standards, how much does history matter?

Cory P. Knobel  
*cknobel@umich.edu*

August 6, 2006

## 1 Introduction

In the study of infrastructure, the issue of standardization is key to gaining an appreciation for the ways in which heterogeneous systems of processes interface to form a larger, working whole. On the technical side, standards work to promote interoperability, linking together various technologies and innovations.<sup>1</sup> Additionally, scholarship has made a strong case that standards are also a field to be studied socially, involving issues of information dispersion or diffusion, enrollment, and communities of practice.<sup>2,3,4</sup> In both cases, attention is paid to the histories of standards emergence. The antecedent events to the formation of a standard are widely considered to be the basis for causal explanations. Further, historical analysis and rich description of the various political, social, and technological events leading to infrastructure are considered by some a foundation for determining a causal path.<sup>5</sup>

In this paper, two accounts of standards emergence are described. Both Paul David and Scott Page use the development of the QWERTY keyboard standard as emblematic of their respective theories in path dependence. The main point on which they differ, however, is the necessity of event ordering in describing that path dependence. It is non-trivial to note that both perspectives are strongly based in the neoclassical economics perspective, which may have significant departure from purely qualitative accounts of historical events, imparting causality through proof. That is to say, economic explanations often imply “it makes sense that things have unfolded this way, because they are axiomatically consistent,” as opposed to narrative accounts of “it seems that event  $X$  precipitated events  $X + n$ .” Page in particular raises interesting and counterintuitive points about situations in which the existence of events matter, but not always their particular order. To be clear from the onset, I am not advancing the idea that historical analysis’ claim that specific events or episodes do not have effect on the subsequent ones at all; however, I make the claim that there may be vignettes or periods in infrastructure stories where tipping points occur, and during those periods, the ordering of events is not the focus of the narrowly defined path dependence.

---

<sup>1</sup>Egyedi, T. (2001) “Infrastructure flexibility created by standardized gateways: the cases of XML and the ISO container.” *Knowledge, Technology, and Policy* 14(3), pp 41-54.

<sup>2</sup>Fomin, V., Keil, T. (2000) “Standardization: bridging the gap between economic and social theory.” *Proceedings of the 21st international conference on information systems*. Queensland, Australia.

<sup>3</sup>Bowker, G., Star, S. (1998) “Building Information Infrastructure for Social Worlds - The Role of Classifications and Standards.” in *Community Computing and Support Systems: Social Interaction in Networked Communities*. Springer-Verlag GmbH. New York, NY.

<sup>4</sup>Wenger, E. (1999) *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press. New York, NY.

<sup>5</sup>Mason, R., McKenney, J., Copeland, D. (1997) “An Historical Method for MIS Research: Steps and Assumptions.” *MIS Quarterly*. September 1997, pp. 307-320.

The functionality of standards in building and maintaining infrastructure are many. Patrick Feng provides five specific purposes for standards-setting.<sup>6</sup> Of the five functions of standards, two are oriented toward technical goals (uniformity in production and compatibility in technologies), two are social (standards as justice and standards as hegemony), and one a hybrid (objectivity in measurement). As I will explain, the interpretation of path dependence is strongly influential in advancing the understanding of standards in both social and technological frames, and by extension, infrastructure.

## 2 Two theories of standardization

### 2.1 Paul David and path dependence

Paul David, Oxford economist and one of the most persuasive advocates of path dependence, has provided ample explanation regarding the standardization of the QWERTY keyboard. He asserts that the history of development, and the specific order of events in which QWERTY was diffused through the population is a necessary condition for its entrenchment as a dominant standard.<sup>7</sup> In the ensuing years, many have reacted to David's assertion that history is the driving force behind the economic development of standards "lock-in" as a concept. In response, he clarifies by defining his terms in a less qualitative way, turning to both positive and negative externality explanations, describing path dependence as an ergodic process.<sup>8</sup>

What, specifically, does this mean with respect to standardization and history? First, let us begin with an explanation of David's theory. Using the QWERTY example, David readily admits that the empirical evidence regarding keyboard layout suggests that QWERTY is not, in fact, the most efficient design. This is in direct conflict with standard economic theory, which assumes that the utility maximization equilibrium is achieved by the most efficient or highest quality solution. The Dvorak Simplified Keyboard (DSK) consistently is shown to be superior in terms of typing speed and user accuracy. Why, then, is it not the standard? David tells a tale based on positive externality and network effects of technology adoption and standardization. With aggressive marketing campaigns, the QWERTY keyboard was introduced to the population through typing competitions showing that they layout was superior with respect to other existing solutions at the time. This, of course, was largely a function of the physical characteristics of the technology. Since manual typewriters were dependent on metal arms moving upward to strike the paper, a fast typist would find that the arms would hit one another and she must slow down in order to work the machine properly - a less than efficient solution, and according to economic theory, a non-equilibrium point. QWERTY, at the time, provided the least mechanical problem, since the layout provided the most distance between arms of the most commonly typed letters, thus minimizing mechanical problems. Further, there were marketing tricks achieved by QWERTY; namely, the word TYPEWRITER could be typed without removing one's fingers from the top row. This gave the impression to buyers that the QWERTY keyboard was the superior product. Now, back to the network effects argument. Once typists were trained in the QWERTY system, a company who had a battery of trained typists would experience standards "lock-in" - a state where the transaction costs of switching to another product are higher than the perceived long-run gains that the new technology would bring. This network effect worked from both sides - the employer and the employee. Employers would experience a dip in productivity during a period of training in another typing system, as well as incur the cost of replacing physical equipment. Job seekers who were called upon to type were most likely to market QWERTY typing skills over other systems, since the companies had invested heavily in QWERTY equipment, thus there was little to no incentive to pursue training on other systems, since they were unlikely to be

---

<sup>6</sup>Feng, P. (2003) "Studying Standardization: A Review of the Literature." *IEEE SIIT2003*, pp. 99-112.

<sup>7</sup>David, P. (1985) "Clio and the Economics of QWERTY." *American Economic Review*. Vol. 75, pp. 332-337.

<sup>8</sup>David, P. (1997) "Path Dependence and the Quest for Historical Economics: One More Chorus of the Ballad of QWERTY." Discussion Papers 020, Oxford University, Economic and Social History.

encountered in the work environment. David argues that the convergence of these externalities, and the diffusion patterns of the QWERTY keyboard form a historical path upon which the standard is based. With respect to defining this convergence as an ergodic process<sup>9</sup>, the strict definition also implies a certain level of technological determinism. This is unsurprising, since modern economics as a field is defined primarily by deterministic equation modeling.

It is clear that David believes that in the matter of standards and infrastructure, history matters. He does admit that there may be other externalities that influence the path dependence of a standard; however, his rhetoric indicates that historical order is the most influential when viewed through the lens of transaction cost economics and a focus on positive externalities.<sup>10</sup>

## 2.2 Scott Page and *phat* dependence

In David 1997, he issues a challenge to the academic community, stating that no one has offered a clear explanation or set of definitions that suggest anything other than the strict view that history matters and that path dependence is non-ergodic. Scott Page - economist, political scientist, and complex systems theorist at the University of Michigan - has taken up this charge and offered an alternative explanation to the strict ordering of historical events in the story of path dependence and standards emergence.<sup>11</sup> While Page does not refute the idea that history matters at all, he takes issue with the idea that all events must happen in a specific order. He defines another possibility, *phat* dependence.<sup>12</sup>

Page makes the distinction between different types of processes. The exemplars chosen by David are typically *Bernoulli* processes - which are by nature ergodic. On the micro-level of interaction, many processes follow a non-ergodic *Polya* process, by which the independence of a long-range set of events is in question. This differentiation brings forth the idea that there may be different levels of analysis that may be applied to history, and that this is significant with respect to interpretation. As Page states, “Even though the Polya Process is only *phat* dependent, this does not imply that the real world situations it has been used to describe do not depend on the order of the path. Those are, of course, empirical questions. Evidence tilts strongly in favor of both types of dependence.”<sup>13</sup>

According to Page, the root causes of path dependence lie not within positive externality; rather, it is the negative externalities that promote the path, and positive externalities only exaggerate the aggregate effects. Negative externalities, in fact, cause historical dependence. Page argues, again through the case of the QWERTY keyboard standard, to suppose that there is another competing design, ZRJSOC. He shows through a series of examples that in a differentiated market (which we shall assume exists, both in the typewriter examples, and in the markets supplying the component systems for building infrastructure), individual consumer choices in adopting standardized technologies may be the source of positive externalities for the choice made, but do not in fact generate negative externalities for substitutable products. That is to say, if one can make multiple choices, choosing a QWERTY device may raise the value of all those owning them, but it does not lower the intrinsic value of the ZRJSOC device. If two technologies are sufficiently competitive, buyer will end up buying both in the long run. This leads us to the fact that negative externalities, which create history dependence (developed on pp 30-31 of Page’s article, but not repeated here) are the precursors of constraints - the harbingers of infrastructural convergence. As Page points out, “Any

---

<sup>9</sup>For clarification, an ergodic process is a positive recurrent aperiodic state of stochastic systems; tending in probability to a limiting form that is independent of the initial conditions. That is to say, it makes little difference what the initial state of a system is. The convergence of probabilities based upon history iteratively narrows the choice set of future interactions, and eventually becomes a single path by which one may travel. Is this not the essence of determinism?

<sup>10</sup>*Ibid.*

<sup>11</sup>Page, S. E. (2005) “An Essay on the Existence and Causes of Path Dependence.” *Santa Fe Institute Working Paper Series*. Santa Fe Institute; Santa Fe, NM.

<sup>12</sup>*Phat* dependence is related to, but not the same as path dependence. The name is a clever rearrangement of the letters in the word *path*, indicating that history matters, but the order may not always.

<sup>13</sup>*Ibid.*

large public decision, be it a prison or university takes up space and requires money. Both create negative externalities with future public projects. Obviously, the more money and space a project demands, the greater its impact on the path. Small projects are less likely to influence the path of history than are large projects. That is not to say that smaller decisions cannot accumulate over time and restrict history to certain paths, but that any big project crowds out other projects.”

### 2.3 Main points of agreement and divergence

David and Page present theories of standards development that are, in fact, highly compatible. More to the point, Page is careful not to directly discredit David’s highly entrenched (though hotly debated) view of path dependence. Both theories claim that history does matter when assigning causality to large scale phenomena - those that are adequately described as Bernoulli processes. They agree that the state dependence of a standard is a matter of the states that came before it; however, the road taken to that path, the process, may have variance. As Page posits, “For obvious reasons, these processes generate history dependent outcomes. the history determines the state and the state in turn determines the distribution over outcomes.”<sup>14</sup> Adhering to the definition of an ergodic process, if the probability distribution over this set of outcomes becomes sufficiently narrow, path dependence is the result. If a wider distribution is observed, a case for path dependence may be mounted.

At its core, I believe the implied and largely unstated difference between the two theories lies in the level of scale in forming an analysis. In the case of QWERTY, David concentrates on standardization as a stationary, ergodic state dependent process. Each preceding episode of entrenchment of the QWERTY design generated strong positive externalities, with the aggregate effect generating negative externalities, not for other designs, but for the industries and firms (both traditional firms, and the job-seeking workforce that must make itself marketable) as a function of limited resources (capital to invest in design-specific equipment, and time and effort for learning a system, respectively.) As Page points out, on the individual level, the advancement of QWERTY did not generate negative externalities for other designs themselves, and consequently, the sequence of adoption of the QWERTY standard over others was relatively inconsequential at the micro-level. It was not until the network effect of QWERTY had reached a sufficiently large level that path dependence was a certainty. This is the foundation of David’s argument, in which the aggregate (and correct at this level of scale) observation that positive externalities drove the closure of the QWERTY standard. David, however, may be erroneously assuming that these same positive externalities were the driving factor in adoption at the individual level. Page makes a convincing case that this is not the entire story. Not surprisingly, this is quite in line with the standard “complex systems” point of view - that micro-level interactions (without negative externalities) lead to macro-structures with emergent properties (clearly identifiable negative externalities.)

## 3 A potential case for further research

How, then, can these concepts of path and path dependence be extended to inform the role of standards development in the field of technological infrastructure? One idea for research can be found in the study of technological innovation as viewed through the patent literature. Most technologies created in the modern era are composite technologies, comprised of patents held by a range of firms. For example, MPEG-2 technology, used in the encoding and decoding of video and audio, is the standard used in creating a number of devices (VCR, DVD, MP3 players, media applications, etc.) This standard, administered by the MPEG-LA group, was defined by identifying the essential non-substitutable patents required to execute the technology. Was the evolution of the MPEG standard path or path dependent? Of particular relevance is that a patent pool was formed around the

---

<sup>14</sup>*Ibid.*

MPEG standard, giving legitimacy to the standard by creating positive externalities for those who hold patents in the pool license portfolio, and negative externalities for those who do not, and for consumers who own technologies that do not incorporate the standard. In this sense, the emergence of the MPEG standard would seem to fit well with David's path dependent theory. Looking at the micro-level history of the standard formation, Page's theory of path dependence may also prove to be informing.

Carl Shapiro hypothesized the existence of patent thickets - areas of high density in patent space where cross-citation is particularly heavy, denoting the inter-relatedness of innovation in a particular area.<sup>15</sup> It has been shown that standards and patent pools tend to form within existing patent thickets.<sup>16</sup> Since a patent pool is dependent upon a standard of technologies against which essentiality may be evaluated objectively, we can turn to the emergence of the patent thicket to see the nature of path dependence in the area of innovation. For example, if an eventual technology that forms a patent pool contains, say, 15 essential patents (which, given the history of pool development, are culled from a larger number of highly related patented technologies, i.e. a thicket. The MPEG patent pool contains 65 patents; however, these were culled from the inspection of over 800 patents submitted for evaluation), does the order in which the last three (or any number. I have chosen three for the sake of argument) are invented matter?

According to Shapiro, the criterion for defining a patent thicket lies only in the recognition of a dense area in the patent space. Research is currently being conducted to measure the minimum density by which one may identify thickets.<sup>17</sup> Assuming that this minimum density can be identified, a standard may be formed around the emerging technology. It does not matter in what order the essential patents were invented and filed, since the formation of the standard is only started when the recognition of this density occurs. In this case, the formation of the standard is a non-ergodic process, as the paths which gave rise to a number of the component technologies are likely not co-dependent. Further, once the standard is developed, the remaining three pieces of technology required to construct the device are likely to be created quickly. Why, one is tempted to ask? Thomas Hughes provides the explanation through his concept of *reverse salience*.<sup>18</sup> Given that the standard exists, and enterprising innovators will be able to clearly see which aspects of the technology standard are unfulfilled, they will fill in the needed gaps, and the order in which these patents are filed may again not matter. This is not to say that several innovators will not create solutions that are redundant or substitutable; however, the formation of a patent pool around a standard disallows the inclusion of substitutable components. Thus, the patent pool itself may be path dependent, but the formation of the standard on which it is based is likely path dependent, as may be the evolution of the composite technology itself.

As it would seem, the process begins as non-ergodic, enters a period of ergodicity around the time of standard-setting, and may or may not enter another period of lower-level ergodicity until convergence, but this latter period does not undo the effects of the middle period. This, of course, is speculation. Research in this area may concentrate on the timelines upon which a particular thicket is identified, essential patents are filed, and a composite technology comes to bear.

---

<sup>15</sup>Shapiro, C. (2003) "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting." *Law and Economics* 0303005, Economics Working Paper Archive EconWPA.

<sup>16</sup>Clarkson, G. (2006) "Sharper Machetes for the Patent Thicket: Objective Criteria for Antitrust Evaluations of Patent Pools." Harvard Business School Working Paper Series.

<sup>17</sup>Clarkson, G., Jian, L., Knobel, C. (working paper) "Measuring Patent Thicket Density: A Network Analytic Approach to Patent Informatics." Accepted to 2005 Conference of the Academy of Management. Submitted to *Strategic Management Journal*.

<sup>18</sup>Hughes, T. P. (1989) "The Evolution of Large Technological Systems." in *The Social Construction of Technological Systems*. Bijker, Hughes, and Pinch (eds.), pp. 49-82

## 4 Conclusion

In the end, it is clear - history does matter. The extent to which it matters, though, is a question of the level of scale at which historical analysis is conducted. It is indeed unnerving, especially to historians, to think that the sequence of events are not the root of causality. This fear is not unfounded, and economic theory supports this principle in many, if not most cases. On counterpoint, once the level of analysis focuses on sufficiently short term, small, or individual histories, Page's concept of Phat Dependence may gain traction. As an analogy, this differentiation between large scale Bernoulli processes and small scale Polya processes may be mirrored by the differences between classical Newtonian physics and the later recognition and adoption of quantum physics. The simplicity of the classical approach holds and is accurate when looking at phenomena at a sufficiently large scale, but these rules do not perform as expected when observing elementary particles. So it may be with path dependence in the economics of standards. Traditional historical analysis is quite accurate in assigning causality through the observation of sequence, but when aggregating micro-level phenomena, such as individual consumer choices, these same assumptions of sequence causality may not be as persuasive. In short, telling the story of standards development, and their eventual role in the construction of infrastructure, may require a blending of both types of dependence, path and phat.

In terms of determinism, historical economics provides a strong argument that once a certain level of enrollment is achieved, the constraints generated by negative externalities may be quite causal in directing development along a certain path. For those who are forward-looking enough, predictions about the eventuality of infrastructural forms may be more accurate, and decisions may be made regarding whether this is a sufficiently optimal outcome, or whether changes need to be made early on, before the onset of "lock-in." Path dependence seems to be a moderately strong statement of determinism, while phat dependence, claiming that outcomes are independent of sequence or initial conditions, points toward undeniably strong deterministic interpretations. It remains to be seen whether this is problematic for scholarly communities; however, it casts a shadow of doubt over those arguments that reject determinism outright.