



Mesmerised by Numbers
Photo: Hsing Wei, Flickr

Counting Contention

Manfred Elfstrom

In the past few years, a growing number of academics and activists have launched projects aimed at counting contention in the realm of Chinese labour. This essay explores the power and limitations of such efforts, detailing the inevitable data problems involved in any quantitative approach to documenting protests in China. It also examines the ethics involved in how we collect such data and the questions we ask of it.

In August 2017, activist Lu Yuyu was sentenced by the Dali City People's Court to four years in prison for 'picking quarrels and stirring up trouble' (*xunxin zishi zui*). Together with his partner Li Tingyu, who was detained but then eventually released, Lu had collected information on thousands of 'mass incidents' (*qunti shijian*). These incidents encompassed a broad range of Chinese contentious politics, from pitched battles between farmers and thugs hired by development companies, to taxi drivers parking *en masse* along roadways in protest against high fuel prices; and from migrant workers threatening to jump from bridges, to retirees gathering to condemn pollution. Li and Lu posted the information they gathered on social media and blogs under the name **Wickedonna** (also known as *feixinwen*, non-news). Their reports usually took the form of short, two- or three-sentence descriptions of confrontations, paired with photos or videos. But the activists also calculated monthly and yearly totals of periods of unrest, accompanied by brief analyses.

It is impossible to know what, exactly, triggered the initial detention of the pair in June 2016. Drawing data of this sort together in one place was clearly unsettling for authorities. But *what* exactly was so troubling about it? One possibility is that making individual stories of conflict so accessible was dangerous. The activist Huang Qi was detained half a year earlier because of his work curating similar tales on his website 64 Tianwang. Most likely, though, the sheer *number* of incidents Lu and Li recorded undercut government narratives related to the creation of a 'harmonious society' (*hexie shehui*) and bringing about the 'great revival of the Chinese people' (*zhonghua minzu de weida fuxing*). After all, the government itself used to sporadically make public annual incident counts, but it stopped doing so in 2005, presumably because the numbers conveyed a negative impression of where the country was headed.

Wickedonna represented a particularly dedicated and courageous effort at counting Chinese unrest. However, a growing number of academics and activists have launched similar projects—albeit mostly from the safety of universities and non-governmental organisations located outside the mainland. These projects include **an extensive dataset** covering all types of disturbances put together by Chen Chih-Jou of the Academia Sinica; my Ash Center colleague **Li Yao’s dataset of all the mass incidents** reported in *Boxun* between 2001 and 2012; the **‘China Environmental Protests & Accidents’** map covering 2005 to present; **China Labour Bulletin’s (CLB) strike map** of workplace unrest from mid-2011 to present; my **own China Strikes dataset** covering the Hu–Wen decade; and **the Global Labour Conflicts collection** hosted by the International Institute of Social History in the Netherlands, which combines the maps from CLB and myself for its China section.

In this essay, I will explore the power and limits of such efforts with regard to Chinese labour issues. I will begin by contrasting the practices of digging deep (qualitative research) versus counting (quantitative research), and posit that both have their virtues. Then, I will detail the inevitable data problems involved in any quantitative approach to documenting protest in China, while arguing that these problems, although serious, should not deter researchers. Finally, I will examine the ethics involved in how we collect such data and the questions we ask of it.

Digging Deep versus Counting

Research about Chinese labour politics to date has tended to involve in-depth fieldwork (digging deep). Scholars have worked on production lines themselves or interviewed strikers in cafes outside factory gates. They have also put in less adventurous,

but sometimes equally demanding, time navigating the Chinese bureaucracy to buttonhole labour officials and trade union leaders. This is as it should be. To a certain extent, you have to physically be ‘there’ to ‘get it’. There are also good reasons to be suspicious of statistics in labour studies. Quantitative data introduces the temptation, often subconsciously, to begin to make the sorts of assumptions about rational individuals and limited options that have led mainstream economics astray. Numbers, moreover, necessarily omit the voices of workers themselves. No econometric take on the industrial revolution, for example, will ever have the humanity of **E.P. Thompson’s exploration** of the many forces that ‘made’ the English working class. And most basically, with regard to China in particular, reliable quantitative data is simply not available for many aspects of the country’s economy and society.

Nevertheless, counting unrest, where it is possible, serves at least two important purposes. First and most importantly, it can give us a sense of the general trends that underlie individual cases and how those trends interact—an imperfect sense to be sure, but a sense all the same. While a close analysis of a single strike at a single factory can challenge our ideas of how things ‘work’ in China and provide us with important clues as to the mechanisms of change at the grassroots level, we should not use a single case—or five or ten or twenty cases—to make claims about how *most* conflicts play out. Labour scholars generally include in their qualitative work various asides about limited external validity, but they nonetheless too often leave the distinct impression that they are really talking about where China *as a whole* is headed, caveats notwithstanding. You simply cannot do this with a country so big and complex.

Second, and this is maybe not so obvious, the process of counting itself is illuminating. Reading strike report after strike report can

help you develop a rough sense of what the norms are for conflicts, something which is hard to come by if you are not yourself involved in activism on the ground. At the very least, you develop a feel for the norms of *reporting* a conflict. Thus, for example, I have noticed that coverage of public bus strikes in Chinese state media tends to take the following form: ‘Today, many people waited for their buses for hours but none came. This reporter went to the bus station to investigate. Drivers were standing around. They had several important complaints. When asked about the issue, the transportation bureau said that they were putting together a team to investigate...’ Then, when you read something that deviates from the expected story arc, it makes you take notice. You know that it is an outlier—and you wonder what has changed.

Counting contention has yielded useful findings in other labour contexts. For example, sociologist **Beverly Silver’s** *Forces of Labor*, which describes how worker militancy chases particular industries around the globe, continues to be a touchstone for many scholars (and has been translated into Chinese, winning it a new circle of readers there). **Edward Shorter and Charles Tilly’s examination** of the size and frequency of strikes in France from approximately the mid-nineteenth to mid-twentieth centuries draws together national politics and shop floor dynamics in an illuminating manner. **Graeme Robertson’s study of Russian labour protests** under Yeltsin and Putin similarly documents fascinating interactions between workers and elite actors in a hybrid regime. Works like these do not rely on perfect data (Silver’s World Labor Group dataset uses articles from *The New York Times* and *The Times* of London to cover unrest the world over), but they are nonetheless revealing. There is no reason this work should not be extended to the world’s second largest economy and largest working class. However, any quantitative approach to China must be balanced by extra close attention to the data

problems and ethical issues involved in how we go about gathering and using numbers. I turn to these next.

Data Problems

If we are to begin counting, we must address the problems with the data available to us upfront. The most obvious problem is with totals. Everyone naturally wants to know how often, and in what numbers, workers are protesting in China. However, it is unlikely that any individual or institution actually has this information—not even the Chinese government. CLB has tried to be clear that their strike map likely only represents a small sample of the full ‘population’ of unrest occurring. Nonetheless, media outlets have repeatedly treated the CLB map as a more or less accurate measure of year-to-year changes in contention. This problem has sometimes extended to scholars. At one point in her **thought-provoking 2016 article** in the *Journal of Asian Studies*, which was the focus of **a discussion** in a previous issue of *Made in China*, Ching Kwan Lee contrasts CLB figures with numbers she obtained from the Shenzhen City Labour Bureau and Ministry of Public Security in the 1990s and early 2000s, concluding that ‘the current period is actually witnessing a decline, not a rise, in strikes’. But surely the government sources Lee cites, while themselves imperfect, convey a more complete picture than CLB. Although her broader point stands—there was a lot of protest before, too, and academics are overstating today’s conflict—the precise comparison made is thus inappropriate.

Other problems relate to more subtle issues of bias. Although a surprising amount of information is reported from poorer and more remote parts of the country, more data will always be available for a place like Guangdong than Gansu. Guangdong, in particular, is home to a relatively open press (especially the Southern Media Group),

scores of bloggers and social media gadflies, and foreign journalists—and it abuts Hong Kong. Whether the imbalance in reports from different places is wildly out of sync with the imbalance in actual instances of unrest is a question that is important but almost impossible to answer. There are also biases in terms of sources. Dissident media outlets are unsurprisingly better at documenting violence against workers by police or thugs, whereas state media makes sure to always describe the steps taken by officials to resolve a dispute. There may also be biases that manifest themselves over time, as censorship tightens and slackens, and news cycles come and go. Statistically, these issues can be dealt with by using only one source for all reports (so that the biases are predictable), by including fixed effects, by zeroing in on dynamics in a particular region, or by seeing what happens when all observations with more than a certain number of incidents are dropped (to ensure that the story is not just related to one of the hotspots). The main thing, though, is to be honest about the data's limitations.

The Questions We Ask

Not all the dangers of counting relate to issues of samples versus totals or bias in terms of sources or cycles of reporting. We should also consider the real-world impact of what we are trying to learn with quantitative data. As with qualitative research, sometimes the easiest questions to answer are not the ones that matter most. Moreover, focussing on certain issues over others means serving certain interests over others. Traditional studies of strikes by labour economists have focussed on the number of incidents, number of participants, or the number of workdays lost. These figures are tracked by many governments (although not China's) and obviously have a big impact on an economy. But why not put more energy into **rigorously**

documenting the different types of worker claims made? Claims may matter more to discussions of class consciousness. When you have thousands of 'dots' representing disputes marked on a map—the default of software programs like the one I use for my *China Strikes* map—it is natural to wonder why they cluster in some places and not in others. However, workers and their allies may not need lessons about what *causes* mobilisation. Instead, they may be more interested in its *consequences*: which tactics are effective and which are not, where in the country breakthroughs are possible, etc. In contrast, state authorities and businesses have an intense interest in understanding the roots of contention: they just want to stop protests! As has been well-documented in the media, big data is now being harnessed to track individuals and groups in China in disturbing ways, such as via the country's proposed new social credit system. The American military, meanwhile, is devoting substantial sums to predicting social upheavals with its Minerva grants. The protest maps assembled by labour researchers are unlikely to be as comprehensive as these projects and so should not raise the same level of concern. But research can always be turned to multiple purposes. It is important for scholars to ponder what purposes their research *most easily* serves. This sort of awareness is the least we owe people like Lu Yuyu and Li Tingyu who have sacrificed their freedom to document what is happening.

Manfred Elfstrom

Manfred Elfstrom is a China Public Policy Postdoctoral Fellow at Harvard University's Ash Center for Democratic Governance and Innovation. He has a doctorate from Cornell University's Department of Government. Before beginning his doctoral studies, he worked for several years with non-profit organizations supporting worker rights in China.