

From Marx to emotions

Lecture given on the occasion of the award of honorary doctorate,
University of Oslo, 2017

Jon Elster

Professor, Columbia University

je70@columbia.edu

My current work is focused on emotions. In this talk I shall sketch the zigzag intellectual and personal journey that, over some fifty years, led me to that topic. The talk will also be a reflection on the development of the social sciences in the same period. As I shall explain, the title, “From Marx to emotions”, covers two very different roads. Each has several lateral roads, and some roads that were initially parallel ended up converging.

I was born into an intellectual left-wing family. My father was, among other things, the nearest the Norwegian Labor Party ever came to having an ideological spokesman. He took a very pragmatic approach, however. He is often cited as having asserted that “Socialism is whatever the Labor Party currently has as its platform”. He never wrote or said that, but the statement is not too far from the way he saw politics. Because of the family context, and other influences as well, I grew up as an unthinking socialist, and never asked any hard questions about why socialism was superior to other economic or political systems.

When I ultimately felt the need to justify my beliefs, I decided to study Marx. I assumed, wrongly as it turned out, that socialism had to be anchored in Marxism. Since I also assumed, only partly wrongly, that Marxism was anchored in Hegelianism, I first undertook a study of Hegel’s *The Phenomenology of Mind*. It was in some ways a rewarding enterprise, but did not provide any foundations for politics. I went on to write a dissertation on Marx, defended in 1972 and published with many changes 13 years later.

To study Marx, I obviously had to understand Marxian economics. To assess that system, I had to assimilate standard, non-Marxist, economic theory and decide which was the more plausible. I discovered that Marxian economics was untenable, because it left no scope for producer or consumer *choice*, except for the minimal choice involved in the equalization of the rate of profit in all sectors. Capitalists could not choose techniques in any meaningful sense, but they could choose to move their capital from a low-profit sector to a high-profit sector until all sectors had the same rate of profit. Marx did not leave any room at all for consumer choice in the determination of relative prices.

Having decided that standard, or neoclassical, economics was superior to Marxian economics, I decided to study the foundations of neoclassical economics, rational choice theory. In a way, this was a mistake, but perhaps a fruitful one. Today, I believe that the idea of *choice* is at the core of social science explanations, but that *rational* choice is only one variety among others. I might not have arrived at that conclusion, however, without passing through a rational choice purgatory. The theory appears to have a compelling combination of clarity, simplicity, and predictive power. In many cases, however, it turns out to be simplistic, and its predictions either indeterminate or falsified. I shall have more to say about these flaws. Yet the point I want to make now is that by virtue of its clarity and explicitness, the theory forces its critics to explain *exactly how* it breaks down, rather than simply dismissing it as ideology or science fiction.

For some years, I certainly displayed some of the traits of the neophyte. I cringe when I reread some of my enthusiastic endorsements of rational choice theory. Yet I soon became aware of its limitations. In its standard form, the theory assumes that the basic explanatory elements, opportunities and desires, are given independently of each other. In many cases, this makes obvious sense. When I go to the market to buy peaches, my choice is *limited* by what's on sale. Within this opportunity set, it is *guided* by my desires. My desires do not shape what's on sale, nor do my opportunities shape my desires.

However, two classical tales provide counterexamples. In the *Odyssey*, Homer showed how Ulysses let his *opportunities be shaped by his desires* when he let himself be tied to the mast when approaching the island of the Sirens. In his fable about the fox and the grapes, Aesop showed how *desires can be shaped by opportunities*. The titles of the first of two books I wrote about rationality, *Ulysses and the Sirens* and *Sour Grapes*, reflect these two non-standard cases.

In my third book on rationality, *Solomonic Judgments*, I considered one of the two cases in which rational choice can break down, by *indeterminacy*, that is, by the failure to yield *unique* predictions. For an example, consider the quandary of two rational agents playing a game of Chicken, and trying to decide whether to swerve or not. Rationality dictates that each should swerve if and only if the other doesn't, but it cannot tell us what they *will* do. At this time, around 1985, I had not fully identified the other case, *irrationality*, that is, the failure to yield *confirmed* predictions. I tended to see irrationality as a residual category, as sand in the machinery of action, rather than as including positive mechanisms of its own, with predictable implications. As I wrote at the time, you can't beat something with nothing, and rational choice theory, however controversial, was certainly something.

I was, however, picking up some positive alternatives to the theory, associated with the names of Daniel Kahneman and Amos Tversky, the creators of behavioral economics. Their main contribution was to identify mechanisms that generate *cold*, that is unmotivated, irrationality. These are, as it were, analogues to optical illusions, such as the stick that looks broken in the water. In one study, subjects were given a description of a young man with long hair and a habit of reading poetry and asked whether they thought it more likely that he was an orchestra violinist or a truck driver. Most said he was more likely to be a violinist, thus ignoring the base rate of the two groups, that is, the absolute number of individuals in each. There are so many more truck drivers than orchestra violinists in the nation, and so much variation among truck drivers, that the poetic young man is in fact more likely to drive a truck.

Since that time, some thirty or forty years ago, such mechanisms of cold irrationality have proliferated like weeds in an empty garden. Hardly a week passes without a new “effect” or “mechanism” appearing on the Internet. The rational actor has seemingly been replaced by a bundle of unrelated mental quirks. As Dagfinn Føllesdal has argued, however, rationality nevertheless has normative priority since human beings *want* to be rational. This desire acts as a permanent centripetal counterforce to the centrifugal mechanisms of irrationality. I should add, though, that the counterforce can itself be undermined by our irrational reluctance to admit mistakes. As La Rochefoucauld said, “Nothing hurts our self-esteem so much as to disapprove of what we once approved.” My own slowness in coming to terms with rational choice theory may owe something to this mechanism.

I remember very well the exchange that opened my eyes to mechanisms of *hot*, that is, motivated, irrationality. In an article from 1980, Richard Thaler illustrated what he called “the endowment effect” by the following example. In a small suburban community, Mr. H. mows his own lawn. His neighbor’s son would mow it for \$12. He would not mow his neighbor’s same-sized lawn for \$20. Thaler argued that this behavior was irrational, or inconsistent. The fact that Mr. H mows his own lawn while his neighbor’s son would mow it for \$12 reveals that he values the time spent mowing the lawn at less than \$12. However, the fact that he would not accept an offer to mow the same-shaped lawn for \$20 reveals that he values the time spent mowing the lawn at more than \$20. The case may be seen as illustrating the difference between out-of-pocket expenses and income foregone. Somehow, the latter seems to have a more shadowy or less motivating existence than the former, even when, as in this case, the opportunity to earn an income by mowing the neighbor’s lawn is *certain*.

One could argue that this behavior simply reflects a general tendency to prefer a bird in the hand over two in the bush, a tendency that is largely justified by the pervasive importance of uncertainty in human affairs. If not rational in a technical sense, the tendency might at least be *adaptive*. In Thaler’s thought experiment, however, there is no reference to uncertainty. More important, the seeming inconsistency has an alternative explanation. As Amos Tversky suggested to me, the lawn-mowing puzzle might be more plausibly explained by the operation of *social norms* rather than by the endowment effect. A resident would not think of mowing his neighbor’s lawn because there is a social norm in suburban communities against an adult doing such tasks for money. *It simply is not done*. In fact, the neighbor’s offer would be seen as an insult. Around the same time, Fredrik Engelstad also urged me to add social norms to my explanatory toolbox, which I did in a book from 1989, *The Cement of Society*.

In that book I also appealed to *emotions* in their role of sustaining social norms. Thus I can trace the first road from Marxism to emotion as follows. The failure of Marxist economic theory led me to study neoclassical economics and its foundation, rational choice theory. The failure of rational choice theory led me to study alternative approaches, such as social norms and their emotional supports.

In itself, the last step only involved a limited set of emotions, such as shame and contempt. However, a few years earlier, I had taken the first tentative steps towards a more general account of emotion, when I co-organized with Klaus Scherer a conference on rationality and emotion at the Maison des Sciences de l’Homme in Paris. Here is an extract from the paper that I wrote up afterwards:

Marx was massively prone to wishful thinking in what he wrote about communist society and the process of getting there. His mind seems to have been shaped by two implicit assumptions: whatever is desirable is feasible, and whatever is desirable and feasible is inevitable. The first assumption shows up in his refusal to consider trade-offs between values, and in his belief that all good things go together. Also, it appears in his neglect of biological and physical constraints, such as inborn genetic differences or the scarcity of natural resources. The second underlies his unwavering belief in the imminent and immanently necessary communist revolution. Whenever the trade cycle went into one of its downward swings, he used to write to Engels that this time it was the beginning of the end.

These remarks, while true, provide only part of the truth. The other side of the coin is that the emotions of rage, indignation, and hope provided the indispensable motivation for Marx's political and theoretical work. They kept him going through years of exile and misery in London, and sustained his enormous scholarly labors no less than his tireless organizational work. [...] The story I have told suggests that the motivation for achievement interferes with the efficacy of achievement; that *the emotions which in some sense provide a meaning and a sense of direction to life, also prevent us from going steadily in that direction.*

The idea stated in the phrase I have italicized has guided much of my research over the last decades. As you will have noted, this path from Marx to emotions is very different from the one I described earlier. The first path involved the study of Marx's *theories*, the second that of his *personality*. When I was reading Marx in the 1970s it did not occur to me to seek the roots of his intellectual flaws in his personality, and certainly not in his emotions. I took his theories at face value, as one should do, and decided that they were for the most part intellectually untenable. I probably also concluded that an element of wishful thinking was at work, but the discovery of the emotional root of that element came later.

I do not want to convey the impression that Marx did not leave any valuable intellectual legacy. He did, and he embodied it. In the *Economic and Philosophic Manuscripts* that he wrote at the age of 26, he argued that the good life for human beings is one of active self-realization. The most poignant statement of this idea is found in the working manuscript *Grundrisse* that he wrote 13 years later:

In the sweat of thy brow shalt thou labor! was Jehovah's curse on Adam. And this is labor for [Adam] Smith, a curse. "Tranquillity" appears as the adequate state, as identical with "freedom" and "happiness." It seems quite far from [Adam Smith's] mind that the individual, "in his normal state of health, strength, activity, skill, facility," also needs a normal portion of work and the *suspension of tranquillity*. Certainly, labor obtains its measure from the outside, through the aim to be attained and the obstacles to be overcome in attaining it. But Smith has no inkling that **this overcoming of obstacles is in itself a liberating activity**. [Labor] becomes attractive work, the individual's self-realization which in no way means that it becomes mere fun, mere amusement, as Fourier with grisette-like naiveté, conceives it. Really free working e.g. composing, is at the same time precisely the most damned seriousness, the most intense exertion.

Marx knew what he was talking about. The first volume of *Das Kapital* is a work written for the happy few, by one of them. It makes no concessions whatsoever to the uneducated reader. Marx assumes that his readers know Latin, Greek and the main European languages. They should be equally well versed in philosophy as in political economy, with a firm grasp of world history and current political affairs. Moreover, they should be able to recognize

literary allusions even in fairly disguised forms. It is a book that stretches the reader's mind to the limits, as it had no doubt stretched its author's capacities. It is, in other words, an extreme feat of self-realization. In the future communist society, everyone would be capable of understanding works of this stature. Indeed, everyone would be capable of *writing* comparable works, and devote most of their time to doing so. To be sure, Marx was not as Utopian as Trotsky, who wrote that in the Communist society "The average human type will rise to the heights of an Aristotle, a Goethe, or a Marx. And above this ridge new peaks will rise." Yet Marx never confronted the practical problems of organizing an industrial society in which, as he wrote, "the free development of each is the condition for the free development of all", as in the relations among musicians in a small jazz orchestra or the crew in a small fishing boat.

The last remarks suggest that the ideal of self-realization should be embedded in that of *community*, or *self-realization with others*. In the *Phenomenology of Mind*, Hegel described what he called the "spiritual animal kingdom" which I interpret as a form of *competitive self-realization*. You have probably come across scholars who refuse to share their unpublished ideas lest someone should steal them and get the credit. As a result, their reputation is enhanced, but the growth of knowledge is stunted. The structure of interaction is that of the Prisoner's Dilemma. By contrast, in a genuine community of scholars the give and take is such that the members may not be able to recall who first came up with an idea. This was, for instance, my relationship to the late G. A. Cohen, author of the best book on Marx ever published and, incidentally, a frequent lecturer at this university.

The book was *Karl Marx's Theory of History: A Defence*, published in 1978. It is part interpretation, part defense. The interpretation is, in my opinion, definitive. The defense, I believe, fails. I shall first briefly explain the issues at stake, and then discuss a general problem that, in my view, is at the root not only of Cohen's failure but of many similarly flawed theories.

From its inception, Marx's writings on history were plagued by a seemingly insoluble contradiction. In his theoretical statements, he affirmed that technology has priority over property relations. The natural interpretation is that technical change is the independent causal variable, and property relations the dependent one. When writing about particular historical episodes, however, Marx acknowledged the obvious facts that property relations can affect technology—just think of the patent system. Although Engels and others tried to reconcile the claims by arguing that Marx was referring only to priority "in the long run", that move effectively killed the theory by making it immune to refutation. Cohen argued that the contradiction disappears if we distinguish between causal priority and explanatory priority. Suppose we can assert as a general law that "When new property relations would cause a higher rate of technical change, then these property relations will appear". Or, more briefly, "If (if A, then B), then A". The patent system, for instance, would emerge if it would increase the rate of technical change.

Cohen referred to such laws as "consequence laws", because they explain the things we want to understand, the explananda, by their consequences, not by their causes. I believe his argument fails, partly because there are no well-established consequence laws, and partly because if there were, their operation would be entirely mysterious. *How* would a need generate its own satisfaction? Also, the conditional "If A, then B" might not reflect

causal consequence at all, but be a mere correlation, due to the conjunction of “A only if C” and “If C then B”. I shall not pursue these issues, which Cohen and I debated for many years, but use it only as an introduction to the more general question of functional explanation in the social sciences.

A functional explanation of behavior, as I shall define it, appeals to consequences of the behavior that are beneficial for someone or something while not being intended to bring about those benefits. While almost invariably untenable, such explanations exercise a strong attraction on the mind, *due to our deep-seated need for meaning*. I shall provide six examples: one each from Marx, Tocqueville, Pierre Bourdieu, and Michel Foucault, and two from evolutionary psychology.

In Volume III of *Das Kapital*, Marx wrote that:

The circumstance that a man without fortune, but possessing energy, solidity and business acumen may become a capitalist [...] is greatly admired by apologists of the capitalist system. Although this circumstance continually brings an unwelcome number of new soldiers of fortune into the field and into competition with the already existing individual capitalists, it also reinforces the *supremacy of capital itself*, expands its base and enables it to recruit ever new forces for itself out of the substratum of society. In a similar way, the circumstance that the Catholic Church in the Middle Ages formed its hierarchy out of the best brains in the land, regardless of their estate, birth and fortune, was *one of the principal means* [Mittel] of consolidating ecclesiastical rule and suppressing the laity.

A *means* must be a means to an *end*. An end has to be the end of an *agent*. A free-floating agentless end is a meaningless idea. “Capital” is not an agent; only individual capitalists are. Although it may well be true that upwards social mobility did benefit the Church and strengthened the capitalist system, these benefits would not explain the mobility. In the case of the Church, mobility was probably a by-product of clerical celibacy. In the case of capitalism, too, upwards mobility was favored by the absence of caste-like dynasties. Yet the demonstration of plausible benefits of mobility for a corporate actor or an economic system tends to produce a mental click, an *aha-experience*, that is easily confused with the click of explanation, or, as Richard Feynman called it, “the pleasure of finding things out”.

The Old Regime and the Revolution by Alexis de Tocqueville is mostly a sober work of history, anchoring events in individuals and their intentions. Yet in one case Tocqueville succumbed to the functionalist temptation by ignoring the possibility of accidental benefits. He observed, correctly, that when the kings granted tax exemptions to the nobles, they removed any occasions on which nobles might make common cause with the bourgeoisie, and asserted, also correctly, that this fragmentation of the elite worked out to the benefit of the kings. Yet he provides no evidence for his further claim that the kings deliberately used tax exemptions to “divide and conquer” the elites. It is more plausible, and in fact more consistent with some of his other analyses, that this was a case of “*tertius gaudens*”, or third party benefits, when an agent C benefits from conflicts between groups A and B that C did not instigate.

In *La Distinction*, Pierre Bourdieu argued that the subtle rules of etiquette among the elite exist to make it more difficult for outsiders to “crash the party” by imitating the rule-governed behavior. There is no doubt that these rules often have the effect of keeping upstarts down, and at least possible that the elite benefits from excluding them, but these

facts or putative facts do not offer an explanation of why they exist. As many self-proletarianized students have discovered, it is very difficult to break into the working class for someone who was not born into it. In Norway in the 1970s, for instance, young Maoists found that making fun of the royal family was a sure way of alienating themselves from the class they were trying to join. Yet nobody has suggested that the norms of the working class exist in order to make it more difficult for outsiders to pass themselves off as workers. The argument makes no more sense for the norms of the elite. Also, note the flexibility of the argument. Does the elite benefit from allowing upwards social mobility, as Marx argued, or from blocking it, as Bourdieu argued? It seems to be a case of heads I win, tails you lose.

In *Surveiller et punir*, Michel Foucault created the paradigm of *oppression without oppressors* that dominates much of the softer social sciences today. As in the following passage, the genre can be identified by the use of verbs and verbal nouns without a subject:

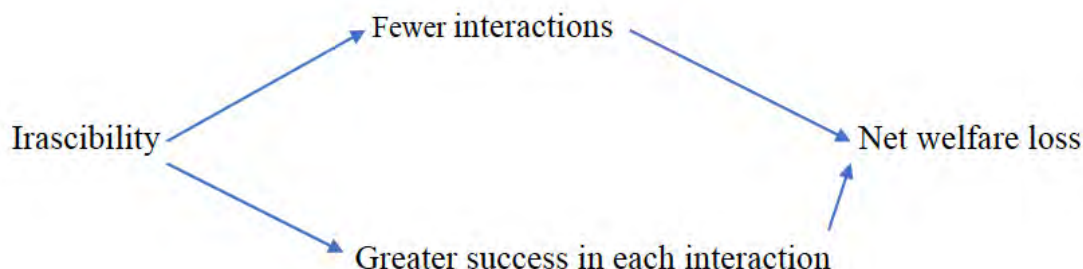
But perhaps one should reverse the problem and ask oneself what is served by the failure of the prison: what is the use [by whom?] of these different phenomena that are continually being criticized; the maintenance [by whom?] of delinquency, the encouragement [by whom?] of recidivism, the transformation [by whom?] of the occasional offender into a habitual delinquent, the organization [by whom?] of a closed milieu of delinquency. [...] Can we not see here a consequence rather than a contradiction? If so, one would be forced to suppose that the prison, and no doubt punishment in general, is not intended [by whom?] to eliminate offences, but rather to distinguish them, to distribute them, to use them; that it is not so much that they render docile those who are liable to transgress the law, but that they tend to assimilate the transgression of the laws in a general tactics of subjection. Penalty would then appear to be a way of handling [by whom?] illegalities, of laying down the limits of tolerance, of giving free rein [by whom?] to some, of putting pressure [by whom?] on others, of excluding [by whom?] a particular section, of making [by whom?] another useful, of neutralizing [by whom?] certain individuals and of profiting [by whom?] from others.

In evolutionary psychology, the demonstration of a net fitness benefit of a given mental disposition strongly suggests that the benefit offers an explanation of the emergence or the persistence of the disposition. There is no question of consequence *laws* à la Cohen, since a mutation that would enhance fitness if it occurred may not occur. Mutations do not occur when needed. Yet if a fitness-increasing mutation happens to occur and is diffused in the population, the existence of the trait for which its codes can be explained by its benefits. In this sense, there are no “non-explanatory benefits” in evolution. However, it is important to verify that the trait provides a *net* benefit, not merely that it has *some* beneficial effects that might be offset by other, negative consequences. I shall give two examples in which scholars failed to consider the issue.

Let me first cite from a book by Robert Frank, *Passions within Reason*, in which he argues that:

Perfectly rational persons with perfect self-control would always seek revenge whenever the future reputational gains outweigh the current costs of taking action. The problem [...] is that the gains from a tough reputation come only in the future while the costs of vengeance-seeking occur now [...] Being predisposed to feel anger when wronged helps solve this impulse-control problem.

There are two things wrong with this argument. First, it ignores two negative effects of anger—the tendency to distort cognition and the tendency to truncate the time horizon of the agent. Rather than solving an impulse-control problem, anger creates one. Second, even if an agent with a tough reputation gets his way in each interaction, he will have fewer of them, because others will learn to avoid him. He may not notice this effect, since he observes only that *anger works*, even if, objectively, *anger works against him*.



A second and more bizarre example is the claim that unipolar depression may have evolved as a bargaining tool, somewhat similar to a labor strike. For instance, an alleged function of postpartum depression is to induce the father to share in raising the child, just as workers go on strike to make employers share the profits. Even assuming that depression can provide short-term benefits, it seems *obvious*—although this is a word to be used with circumspection—that the negative long-term consequences will be larger.

One reason I have spent some time on these functional pseudo-explanations is that I believe there is room and need for a new social-scientific sub-discipline, whose task would be to provide *explanations of bad explanations*. It would be an empirical cousin of the philosophical sub-discipline of *bullshittology*, which is oriented towards conceptualizing the explananda rather than explaining them. I cannot stress enough that such analyses would not be intended to *refute* the explanations. This would amount to “the genetic fallacy”. Refutations must follow standard methodological procedures. Yet I believe that the sheer mass of substandard social science—what I have called soft and hard obscurantism—calls for an explanation. I do not exempt my own writings. In my most recent book, the revised edition of *Explaining Social Behavior*, there is an entry in the index, “mistakes by the author of the present book” that refers to some of my own bad explanations. I shall return to some of them later. In that book, I also propose a number of psychological and sociological mechanisms that may help explaining the amazing persistence of bad social science.

I shall now turn to constitution making, which is my main current research interest. With one exception, which I shall mention, Marx does not have much to say on that topic. What he has to say, however, is directly relevant to the role of emotion in constitution making. But before I get to that issue I shall retrace my steps to display some of my mistakes. As you will see, my work on constitution making went through three stages, as I discovered the complex motivations of constitution makers. In my first approach, I assumed that framers were motivated exclusively by *reason*, that is, by the desire to promote the long-term public interest. Ten years later, as I moved from normative assumptions to actual cases, I realized the fact, which should have been obvious, that they may

also be motivated by personal or group *interest*. Ten years later, I added *emotions* to the repertoire.

Exactly 40 years ago, in an article on “Ulysses and the Sirens” that later became a chapter in the book with the same title, I argued that “the constituent assemblies of the last two hundred years [...] are the closest analogy to the state of mind of Ulysses” when he approached the island of the Sirens. Implicitly, this amounted to stating that framers are in a cool, non-emotional state and take measures to prevent themselves from yielding to their future, predictable, emotional impulses. Although I did note that the idea of a society “binding itself” was a controversial one, I did not dwell on this disanalogy between the individual and the collective case. I was too caught up in the analogy to question it.

My proposition was not a novel one. Spinoza used the Ulysses metaphor to make a similar point. More recently, Friedrich Hayek assimilated constitutions to chains that Peter when sober imposes on Peter when drunk. Yet even though it was not novel, for some reason my argument became quite influential. I now believe that it was profoundly mistaken.

Let me begin with the least serious mistake. It was pointed out to me by my friend and mentor Jens-Arup Seip, who in a characteristically cutting phrase dismissed the idea as nonsense. “In politics, he said, people never try to bind themselves, only to bind others”. Although largely valid, this criticism goes too far. When Tony Blair’s first government came to office in 1997, the finance minister Gordon Brown imposed the independence of the Bank of England, thus *depriving himself* of the politically useful tool of setting interest rates. Yet there is no doubt that many constitutions have been enacted by a majority to bind a present minority or a future majority, a trick that is possible because constitutions are almost always adopted by a simple majority, while amending them often requires a qualified majority.

Another mistake I made in that article was the claim that “the system of periodic elections can be interpreted [...] as the electorate’s method of binding itself and of protecting itself against its own impulsiveness.” Suppose it is true, as it may well be, that the people are better off if they are prevented from recalling representatives at will. Yet that fact by itself has no explanatory power, and certainly does not imply that the people acted to bind themselves. The explanation of periodical elections in terms of the benefits for the people is just a piece of unsupported and in fact implausible functional explanation. It seems much more likely that periodical elections were imposed by *politicians*, whether to protect the people against their own impulsiveness or to promote their personal interest in holding office. We must look at the *intentions* of those who imposed periodical elections, not at the *consequences* of that system.

The most serious mistake was more fundamental. Before explaining it, I shall say a few words about how I came to take a more serious interest in constitution making, compared to my simplistic speculations in the 1977 paper. In early 1990, I was asked to contribute a paper to a conference comparing the Federal Convention that adopted the American constitution in 1787, and the first French constituent assembly of 1789–91. Since I knew nothing about the topic, I accepted, and wrote a paper that was later published under the title “Arguing and bargaining in two constituent assemblies”. While I was preparing the paper for the conference, an extraordinary coincidence occurred. In the spring of 1990, the proto-constitutional Round Table Talks in Poland were taking place. The process was sus-

pended in a force field defined by the risk of a Russian intervention if the regime made too many concessions, and of a general strike if they made too few. While following that process day by day, I was also reading about how the French framers of 1789 were suspended in a force field defined by the king's soldiers and the crowds in Paris. As the title of my published paper indicates, these public confrontations made me understand that constitution making was not simply a matter of calm deliberation in what Jürgen Habermas called “the ideal speech situation”, but also involved threat-based bargaining. As a matter of fact, the American framers, too, engaged in this double process of arguing and bargaining. The idea of collective self-binding against future impulsiveness seemed largely irrelevant.

To put it differently, I was led to enlarge the set of motivations of the framers. They were concerned not only with *reason*, that is, rational deliberation over the long-term public interest, but also with *private interests* and *group interests*, most strikingly illustrated in the bargaining over slavery at the Federal Convention. Yet even this enlargement was too limited, since it only marginally included *emotions* as a motivational factor of the framers. In my published paper, I referred extensively to emotions as the motivation of the future political agents whose behavior would be regulated by the constitution. When I mentioned the emotions of the framers themselves, it was only in the special and narrow context of emotions generated within the assemblies, such as the pride and vanity of the speakers. For all practical purposes I shared the common assumption of most writers on constitution making, that framers are motivated sometimes by ideas, sometimes by interest, and by little else.

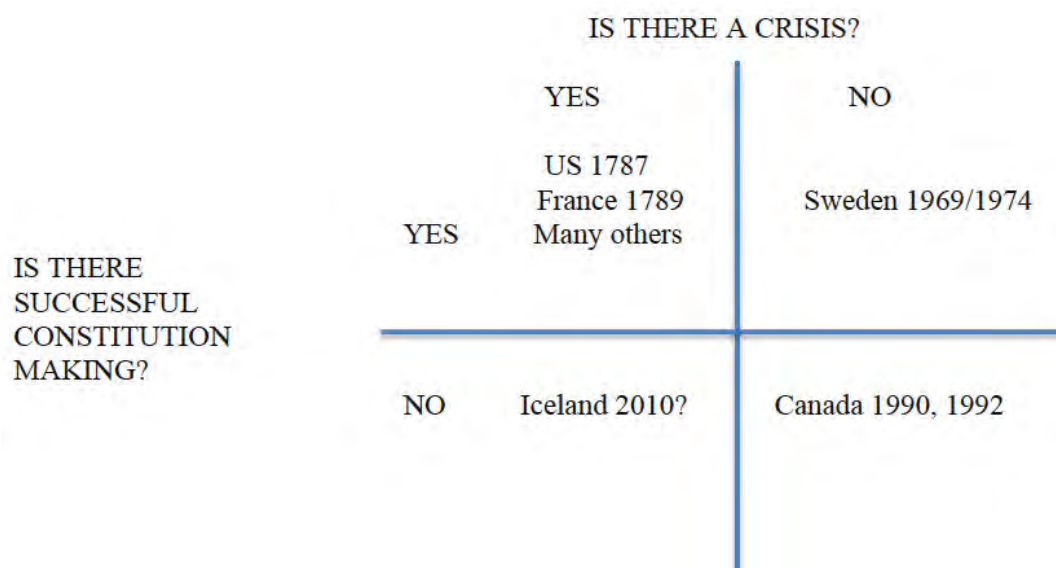
In the 1990s I continued to follow and write about constitution making in Eastern Europe, mostly with few references to emotions. In the same period, I did substantive work on the emotions, leading to the publication in 1999 of a book, *Alchemies of the Mind*, subtitled “Rationality and the emotions”. In that book, there are no references to constitution making. I was proceeding on two tracks, which did not seem to converge.

The convergence occurred when I gave a series of lectures on the French constituent assembly, and had to discuss the momentous decision of August 4 1789, when the framers pretty much abolished feudalism overnight when news reached Versailles in late July that peasants were attacking castles in the countryside. Property-owning deputies abolished not only feudal dues, but also the many privileges of their towns and provinces. The exact motivational mix is unclear. Tocqueville wrote that “the event was the combined product, in proportions impossible to measure, of fear and enthusiasm”. Other historians mostly agree, although some put more weight on fear and others on enthusiasm. What seems clear, from all accounts by participants, observers, and historians is that the assembly experienced an emotional riot. On this occasion and on others, the framers displayed all the characteristic signs of emotional motivation. In *The 18 Brumaire of Louis Bonaparte*, Marx wrote that in the bourgeois revolutions of the 18th century “ecstasy was the everyday spirit”. A commentator wrote that the framers of 1789 were “drunk with disinterestedness”. Their behavior matched the definition of enthusiasm in the Oxford English Dictionary: “Rapturous intensity of feeling in favor of a principle or cause.; passionate eagerness in any pursuit, proceeding from an intense conviction of the worthiness of the object”.

When I discovered the role of enthusiasm in other constitution making processes, including that of Norway in 1814, I looked into the scholarly literature to see what psycho-

logists had to say about this emotion. A long search in handbooks, textbooks, and Google Scholar yielded literally nothing. This fact reinforced my long-standing conviction that academic psychology has become the prisoner of its methodology, like the man looking for his lost key under the lamppost because that was the only place there was light. I shall not pursue this question, however.

The more constitution-making processes I looked at, the more I was struck by the fact that they tend to occur in a time of economic, social, or political *crisis*, with concomitant strong emotions. Although I restricted my work to Western societies, I believe the following diagram is representative of other parts of the world as well.



The “many others” include five French constituent assemblies (1795, 1814, 1848, 1852, 1946), three German assemblies (1848, 1919, 1949), two Nordic assemblies (Norway 1814, Iceland 2010), six post-Communist assemblies (Poland 1989, Hungary 1989–90, Bulgaria 1991, Romania 1991, the Czech Republic 1992, Slovakia 1992), and sundry others (Poland 1921, Spain 1931). In addition, I include proto- or quasi-constituent assemblies, such as the Continental Congress, the Round Table Talks in Eastern Europe and the Consultative Constitutional Committee that, in spite of its lack of formal powers of decision, had a non-negligible influence on the 1958 French Constitution. Many of the American state constitutional conventions were also made in times of high crisis.

On this background, I shall briefly state six propositions:

1. Constitutions tend to be made in times of crisis
2. Crises tend to go together with strong passions
3. Strong passions tend to undermine rational belief formation
4. Only strong passions can generate the political will needed for constitution making
5. Therefore, constitution making tends to be flawed
6. Therefore, constitutions tend (weakly) to be flawed.

Jointly, these propositions imply that we have to turn Hayek upside down. In actual constitution making, framers behave like *Peter when drunk legislating for Peter when sober*.

The first three propositions are, I now believe, relatively uncontroversial. The fourth proposition asserts that *only* the strong emotions generated by a crisis can motivate people to ignore their petty, partisan, and short-term interest in favor of the public good. Let me illustrate it by two statements from the American and French revolutionaries in the 1780s:

All the existing [state] constitutions were formed in the midst of a danger which [...] *stifled the ordinary diversity of opinions* on great national questions (*The Federalist Papers*, No. 49; my italics).

“Anarchy is a frightening yet necessary passage, and the only moment one can establish a new order of things. It is not in calm times that one can take uniform measures” (Clermont-Tonnerre).

The fifth proposition is only a logical implication of the two previous ones. I shall not comment on the sixth proposition, except to say that many constitutions are flawed because of the *haste* in which they were adopted. Sometimes the haste has been due to external circumstances, sometimes to emotionally induced urgency.

I shall conclude, as I began, with Marx. How does my argument relate to Marx’s theory of revolution? My six-point model applies to counterrevolutionary constitutions as well as to revolutionary ones. In my opinion, the American constitution was a counterrevolutionary document, motivated by the elite’s irrational fear of excessive democracy in the individual states. Without the emotions triggered by Shays’ rebellion in 1786–87, the seaboard merchants and slaveholders would not have been able to overcome their differences and unite against the turbulent backcountry. This is a controversial interpretation, but less so than it was a few years ago. Similarly, Tocqueville argued that the crushing of the insurrection of the French workers in June 1848 was possible only because the threat to the bourgeoisie was so acute. “If the rebellion had been less radical and seemed less fierce, he wrote, probably most of the bourgeoisie would have stayed at home”. By contrast, Marx did not propose a general theory of political change that would include both revolution and counterrevolution as special cases.

As I noted, Marx commented on the emotional nature—what he called “the everyday ecstasy”—of the bourgeois revolutionaries of the 18th century. He also asserted that these emotionally based revolutions were *short-lived*. This observation is consistent not only with the findings of modern psychology, but also with a comment by George Washington in a letter from 1782, when the war against Great Britain was as good as won: “I must confess that I am not at all astonished at the failure of your plan [to arm the slaves in the fight against the British]. That Spirit of Freedom, which at the commencement of this contest *would have sacrificed everything* to the attainment of its object, has long since subsided and *every selfish Passion* has taken its place”.

In the same work, Marx went on to contrast bourgeois revolutions with proletarian ones, in a passage as famous and beautiful as it is incomprehensible:

On the other hand, proletarian revolutions [...] criticize themselves constantly, interrupt themselves continually in their own course, come back to the apparently accomplished in order to begin it afresh, deride with unmerciful thoroughness the inadequacies, weaknesses and paltriness of their first attempts, seem to throw down their adversary only in order that he may draw new strength from the earth

and rise before them again even more gigantic, recoil ever and anon from the indefinite prodigiousness of their own aims, until a situation has been created which makes all turning back impossible, and the conditions themselves cry out: Hic Rhodus, hic salta (Here is Rhodes/the Rose, here jump/dance).

Recall now Proposition (4): Only strong passions can generate the political will needed for constitution making. For his part, Marx wrote in 1846 that “the revolution is necessary [...] not only because the ruling class cannot be overthrown in any other way, but also because the class overthrowing it can only in a revolution succeed in ridding itself of all the muck of ages and become fitted to found society anew”. The two ideas have in common an apparent denial of the possibility of a gradual, non-violent, but ultimately radical change. In my model, this implication would follow only on the assumption that all radical change must take the form of constitution making. The existence of Great Britain, which does not have a written constitution, but has changed as much over the last centuries as any country that has one, invalidates this assumption. In 1872, Marx, too, cited Great Britain when he came to consider the possibility of a non-violent transition to socialism: “We know that heed must be paid to the institutions, customs, and traditions of the various countries, and we do not deny that there are countries, such as America and England [...], where the workers may attain their goal by peaceful means”. It seems, in fact, that in Britain, unwritten constitutional norms, enforced by naming, blaming, and shaming, limit self-serving behaviors that in other countries are blocked by a written constitution.

Today, however, these concerns need to be updated. Consider the main problem that confronts humanity: global warming. Assume, more or less accurately, that there exist known solutions to this problem, such as a universal carbon tax. However, any practical implementation of this idea will hurt some countries more than others, and will seem unfair to some countries while fair to others. Proposition (4) implies that no agreed-upon solution will emerge until the situation is so bad that countries and their citizens are willing to put aside national and short-term interests and adopt an enforceable global carbon constitution. The emotion sustaining it might be fear of imminent catastrophe or anger against the political elites. Let us hope that by that time it is not too late.