It seems to me that preserving and strengthening the set of institutions that are so remarkable in human history—our core institutions—is our best hope for ensuring future invention, cooperation, and problem-solving.

Ronald K. Mitchell | October 14, 2016

At a recent meeting of the Society for Business Ethics, I met an ethics panel member who commented on the extra efforts that St. Joseph Health, a very large healthcare organization, was engaging to accomplish its mission: “To extend the healing ministry of Jesus in the tradition of the Sisters of St. Joseph of Orange by continually improving the health and quality of life of people in the communities we serve.” Within their 25,000-person system of 18 non-profit hospitals, they are seeking to unify their efforts by requiring the senior leadership to participate in a four-year organizational formation program and a mission-based immersion experience to internalize and create a values-based culture.
However, thinking further about what I learned from this ethics panel member, it occurred to me that more is needed to accomplish ambitious society-improving goals: St. Joseph Health’s organizational undertaking would not be effective without the support of many core institutions—family, education, ethics, civic virtue, faith, and so on. These institutions provide the framework for cooperative action that benefits individuals, organizations, and society. Thus, when individuals or organizations aspire to affect the good of society, it also is important that they take into account (1) the major social forces that create, maintain, or disrupt core institutions, and (2) the major social forces that influence the workings of the cooperative action that this work entails. I believe that these requirements apply generally. In the case of the Wheatley Institution, these forces must be considered when we undertake our purpose: to lift society by preserving and strengthening its core institutions.

Responding to the first set of forces requires a better understanding of “institutional work.” Responding to the second, requires a better understanding of a rather obscure, but important idea – that of “collectively stable [relationship] strategies” that are pursued by human beings who must live together in a world substantively shaped by scarcity. In this Fellow’s Note, I bring together some of the ideas that define the workings of each of these two forces to explain why, in a competitive society, we can hope—as do St. Joseph Health and the Wheatley Institution—to contribute to the good of society.

**Institutional Work**

The institutions of society exist because people “fix” or “institute” habits of thought and action that, though they may originally be subjective in origin, eventually come to be commonly accepted, forming the institutions of a society. Examples of institutions in Western society include capitalism, a civil society, democracy,
and disruption of institutions. Noted sociologist Émile Durkheim explained that because institutions set expectations for thought and action, they provide the “noncontractual elements of contract”\textsuperscript{[1]}—the social support system that makes contracting possible. More recently, institutions have been viewed as prescriptions for standardized, sequenced human interaction—prescriptions that can be created, maintained, and sometimes disrupted. Thus, institutional work can be defined as the creation, maintenance, and disruption of institutions.\textsuperscript{[2]}

Specifically, institutional creation work includes (1) political work, such as advocacy, law, and rights vesting; (2) belief-system reconfiguration work, which involves constructing networks and identities, or challenging and changing norms; and (3) meaning-system alteration work, consisting of activities such as educating, theorizing, and mimicry.\textsuperscript{2} Maintenance work includes (1) rule-system adherence work, such as enabling, policing, and deterring aberrant behaviors; and/or (2) belief system and norm reinforcement work, which may entail valorizing and demonizing, mythologizing, embedding and routinizing.\textsuperscript{2} Finally, institutional disruption work primarily is deinstitutionalization work such as disconnecting rewards and sanctions, disassociating moral foundations, and undermining assumptions and beliefs.\textsuperscript{2}

Many social science scholars have suggested that institutions so profoundly influence behavior that they form a kind of “iron cage” within which members of a society live from day-to-day.\textsuperscript{[3]} In a world of relative scarcity, where there is constant competition for access to physical, material, and social power, competition for control of the iron cage of institutionalization is expected. And here the age-old juxtaposition of cooperation versus defection (others versus self) confronts us. Therefore, if we are to lift society by preserving and strengthening its core institutions, we must learn how better to harness the forces that lead to preserving and strengthening cooperative action.
We must learn how better to harness the forces that lead to preserving and strengthening cooperative action.

Collectively Stable Strategies
Author Richard Dawkins\(^4\) suggested one possible chapter in the story of developing the concept of collectively stable strategies as follows. The story begins with Robert Axelrod, an American political scientist, who wanted to determine how cooperation evolves. Axelrod considered the many strategies set forth by game theory experts, then set up a competition to see which strategy would work best in the simple but insight-producing game: the Prisoners’ Dilemma\(^5\). Axelrod’s tournament included 62 strategies (plus 1 all-random condition). Each strategy was played repetitively against all others using computer simulations. Axelrod found that when the Prisoners’ Dilemma game is played over and over in this manner, self-interest-based strategies (defection-based) are eliminated, and cooperative strategies triumph. The most interesting part of his finding is the nature of the winning cooperative strategy. In Axelrod’s parlance, the winning strategy in his tournament had the following three attributes: players were “nice”; players were “forgiving”; and players were “non-envious.”\(^6\)

So let’s define “nice,” “forgiving,” and “non-envious” in the context of institutional work. Nice players begin by offering cooperation. Forgiving players make the choice to cooperate when in the prior
move the other player chose was non-cooperative (defection, in
game terms). Non-envious means that the nice and forgiving actions
are adhered to, regardless of the prior-round winnings of the
opponent. Sticking with these actions circumvents the virtually
ubiquitous human tendency of reacting enviously to another
player’s success, then only to end up with a worse outcome. In the
case of the creation, maintenance, and destruction of institutions
amid the competitive motivations that spring from scarcity, what
does it mean to be nice, forgiving and non-envious when such
actions concern institutional work?

The key idea is that the foregoing results are shown to be
possible over time, that is, under conditions of repetition with no end
point specified. Hence, where iterations have a set limit, and where
that limit is known, then the temptation to defect can become
overwhelming and thereby destructive. Also, when urgency
constrains the nice-plus-forgiving strategy, the institutional work of
disruption can be expected. And this outcome can occur especially
when envy becomes the engine of institutional work. So when I
refer to collectively stable strategies in the context of this Note, I am
referring to strategies that preserve core institutions over time.

We will be
able to tell by
the nature of
the
institutional
work the
expectation of
cooperative
outcomes.

Good Institutions
Harvard historian Niall Ferguson[2] has argued
that truly good sets of institutions are hard to
achieve, while bad institutions are easy to get
stuck in—which is why, he asserts, most
societies have been poor, illiterate, unhealthy
and violent for most of our history. He argues
that the human progress we presently enjoy
results from a set of core institutions developed
in recent history. It seems to me, that preserving
and strengthening the set of institutions that are
so remarkable in human history—our core
institutions—is our best hope for ensuring future invention,
cooperation, and problem-solving. In other words, human progress depends on our institutional maintenance work.

What does institutional maintenance work look like? When we engage in institutional maintenance work, we enable creativity and self-determination, respond to anomalies and deter harmful behaviors. We support belief systems that bind together sets of core institutions (e.g., the US Constitution Bill of Rights), and strengthen valued norms—especially norms that involve being nice, forgiving, and non-envious. Our reinforcement work may include valorizing (or demonizing), mythologizing, embedding and routinizing. Such work might involve celebrating human generosity, honoring forgiveness through art, song, and story, and applauding social routines that bolster giving and generosity. By our engaging in institutional maintenance work, even in the face of scarcity and the ensuing competition, that the core institutions of society will be preserved and strengthened.

How can we tell if an institution helps or hinders cooperation? Ferguson asserts that good (i.e., core) institutions incentivize people to do good things. I therefore suggest that we will be able to tell by the nature of the institutional work the expectation of cooperative outcomes. We can ask: Does an institution itself, and the work of maintaining it, produce behavior that is nice, forgiving, and non-envious?

Looking at one present set of core institutions – one which includes capitalism, a civil society, democracy and rule of law – Ferguson warns of a “great degeneration” underway in Western society, with core institutions increasingly in danger of disruption. I contend, however, that a “great REgeneration” also is possible. I suggest that we can learn from game theory that such a regeneration should involve the institutional work of maintenance that leads with cooperation, forgives
defections, and establishes a reservoir of goodwill from which all may draw.

Because the core institutions of society continue to be preserved and strengthened by nice, forgiving, and non-envious institutional maintenance work, I believe that St. Joseph Health—and perhaps all of us—can dare to accomplish an inspiring mission ... and why we can continue to hope for the good in society.

[5] In the traditional version of the game, the police have arrested two suspects and are interrogating them in separate rooms. Each can either confess, thereby implicating the other, or keep silent. No matter what the other suspect does, each can improve his own position by confessing. If the other confesses, then one had better do the same to avoid the especially harsh sentence that awaits a recalcitrant holdout. If the other keeps silent, then one can obtain the favorable treatment accorded a state’s witness by confessing. Thus, confession is the dominant strategy for each. But when both confess, the outcome is worse for both than when both keep silent. The concept of the prisoners’ dilemma was developed by RAND Corporation scientists Merrill Flood and Melvin Dresher and was formalized by Albert W. Tucker, a Princeton mathematician. - The Concise Library of Economics
[6] Since this tournament, subsequent analyses have explored additional assumptions with results varying somewhat according to these assumptions. However, Axelrod’s basic finding has been foundational to the work in better understanding the evolution of human cooperation.