J Value Inquiry
DOI 10.1007/s10790-012-9352-y

BOOK REVIEW

T. Shelling. *Strategies of Commitment and Other Essays*

Cambridge, Mass: Harvard University Press, 2006, 360 pp.
ISBN 13: 978-0-674-01929-4. $39.95 Hb.
ISBN 13: 978-0-674-02967-7. $19.95 Pb

Iskra Fileva

© Springer Science+Business Media Dordrecht 2012

This essay collection comes from an economist and so is not, strictly speaking, a philosophy book. But the author is an outstanding economist, and *Strategies of Commitment* is a very philosophical book. All pieces in it are remarkably lucid and truly engaging; each contains an original idea, and some read as series of insights; many offer examples of the kind that works effectively in a variety of settings – a conference talk, a journal article, a university class or even at the dinner table. Topics covered include old as well as relatively recent interests of the author: commitment as it figures in the contexts of threats, promises, bargaining and self-command, the dynamics of social segregation, arms control and climate change. The book will be of interest to philosophers of economics, rational choice theory, game theory, ethics, political theory and environmental philosophy.

Strategies of Commitment contains a total of nineteen essays grouped in eight chapters. Two chapters will serve as focal points of this discussion: the opening chapter which gives title to the collection and section number three under the heading, "Commitment as Self-Command."

What is "commitment"? The term "commitment" is likely to trigger two groups of associations in the mind of a philosopher. Associations from one group have their origin in a Kantian understanding of commitment traceable back to Plato and recently revived by neo-Kantians. On that understanding, to make a commitment is to bind oneself with reasons, reasons to pursue particular courses of action and avoid others. If one fails a commitment, on this view, one fails as a moral being and possibly, echoing more recent versions of the account, scathes one's own practical identity. The other chain of associations begins with Hobbes's idea that commitments should be made and kept for self-interested reasons. Keeping a commitment is in one's own interest, on Hobbes's reckoning, just in case a strong

I. Fileva (✉)
University of North Carolina, Caldwell Hall, CB #3125, 240 East Cameron Street, Chapel Hill, NC 27599-3125, USA
e-mail: ifileva@email.unc.edu

Published online: 28 November 2012

coercive power exists such as makes it imprudent or – what is the same for Hobbes – unreasonable for either side to fail a commitment. It may seem that these two standard construals of commitment – one normative and the other one instrumental – are polar opposites and exhaust the theoretical possibilities available. But in fact, there is a sense in which the two conceptions are closely aligned and the pair could be contrasted with at least one alternative, Schelling's conception.

The two interpretations of commitment sketched are alike in that both can be read as possible answers to the same guiding questions, "Should I keep a promise made?" and, "If so, why?" Schelling's concern is somewhat different. The problem for him is not whether one should uphold every commitment made and if so, whether the "should" is a normative or an instrumental one. The question for Schelling is, just what kind of behavior is the behavior we refer to when we speak of "making a commitment"? What is the goal of this sort of behavior? How is the goal accomplished?

Schelling's reply is that the kind of behavior properly described as creating or communicating a commitment belongs to a class of behaviors deemed strategic. A behavioral propensity is strategic just in case, "it influences others by affecting their expectations"(p. 21). Commitment is a type of behavior whose constitutive aim is the aim of influencing other people's expectations regarding what one is about to do. This definition has a number of consequences. First, on an account premised upon it, to fail in the attempt to make and communicate a commitment is not, as one would ordinarily suppose, the same thing as failing to uphold a commitment made. The reason is that success and failure are relative to the main purpose of an activity and that purpose here is to influence other people's expectations and other people's behavior. Consider a brief illustration. Suppose that Fred promises Jane to bring a newspaper on the way home so that Jane does not need to go out. His purpose in giving this promise is to stop Jane from going out, maybe, he knows she is feeling unwell but also badly wants a newspaper. Now suppose further Jane thinks Fred is absent-minded and not reliable generally, and imagine she goes out and buys a newspaper despite Fred's promise. However, today Fred doesn't forget what he has said and does bring the newspaper. In so doing he keeps his promise but fails to influence Jane's expectations in the desired way. The strategy he deploys for keeping Jane home fails, his having kept his promise notwithstanding. The second point to note is that no normative requirement is encapsulated in the above definition. Fred can try to get Jane to execute a sinister plan by promising help. This promise of his will be, in all relevant regards, just like the promise to bring Jane a newspaper. Whether or not Fred succeeds in the strategic behavior under consideration depends solely on whether or not his promise has the intended effect. The morality or immorality of the reason he has for giving that promise has no bearing on this issue.

Schelling's conception is clearly not normative in any recognizable sense. But it would be a mistake to conclude from this that it is therefore instrumental. Schelling's account of commitments is not instrumental if by that one means that commitments should be made and kept for instrumental reasons. His account allows for cases in which people make or keep a commitment for strictly moral reasons, and it does not exclude from the class of commitments behaviors motivated

differently. The main question he sets out to answer is not whether and for what reason someone should uphold a commitment made, but when and under what circumstances one person has a good reason to believe that another person will. If one person succeeds in presenting such a reason, she influences the other person's expectations of what she will do and from here, that person's future behavior. Her strategy succeeds.

When does someone have such a reason and what can she do to procure one? This is what the rest of the chapter is about. Schelling's answers to these questions is illuminating and opens up potentially fruitful lines of inquiry. In laying them down, it will be helpful to take Hobbes's reply to the same question as a point of reference: on Schelling's understanding, the existence of a coercive power is not a prerequisite of a commitment's credibility. The reason it is not is that even in the absence of such a power, there are multiple ways in which we can relinquish control over our own actions. We can make it physically, locationally, legally, and in other ways impossible or else extremely costly or inconvenient for ourselves to do something different from what we say we would. We can resort to this capacity in attempting to influence other people's expectations. For example, if Gail wants to prove to Henry that she is not going to harm him, there may be no greater power to persuade him that she will do as she says. But she has the power to disarm herself and so credibly in Henry's eyes commit herself to not harming him. Similarly, if she wants to persuade him that she is not going to read a letter she just received, she can throw the letter in the fireplace. If she wants him to know that her army will fight till the end, she can burn the bridges behind her soldiers and leave them with no way to retreat.

Examples can be multiplied. The last one mentioned takes us back to a story told by Xenophon, whom Schelling credits with an early insight into these matters. The reference appears in the opening passage of the book. That passage conveys well the flavor of the book and merits quotation, "I wrote about commitment almost fifty years ago (Schelling 1960, chap. 2), and some colleagues have conjectured that I originated the concept. That pleases me, but I decline. I was scooped by at least 2400 years. When Xenophon pursued by Persians, halted against an impassable ravine, one of his generals expressed alarm that they would have no escape. Xenophon reassured him: 'As for the argument that …we are putting a difficult ravine in our rear just when we are going to fight … is not this really something that we ought to jump at? I should like the enemy to think it easy going in every direction for him to retreat. But we ought to learn from our very position that there is no safety for us except in victory'" (p. 1).

Schelling says in the preface that in the 1970s he was invited to join a committee of the National Academy of Sciences on Substance Abuse and Habitual Behavior. The committee acquainted him with professionals concerned with heroin, tobacco, alcohol, marijuana, gambling, and eating disorders. A result of this experience was a new rubric in his theory of commitment, one which concerns self-relations, "I had earlier considered commitment as central to influencing the behavior of others," writes Schelling apropos, "it dawned on me that people attempting to control their own behavior often appeared to succeed when they managed to commit themselves to a regime of abstention or performance, treating themselves the way they might treat someone else" (p. ix).

The thought here is that compulsive gamblers, over-eaters, smokers, drinkers, heroin addicts and irascible persons often are people who cannot very well control their own behavior. They employ different strategies *vis-à-vis* themselves, strategies resembling those we employ in attempting to influence others. In describing the general features of this type of case, Schelling appeals to a metaphor of multiple selves: people who try to exercise self-command appear to have not one self but two – a wayward self and a straight self, whereby the straight self attempts to bring the wayward self under control. The expression is meant to capture the idea that a person's future preferences might be different, in some ways radically different, from his present values and preferences. It is as though he can anticipate becoming a different person in the future altogether, someone who would elect to do things which he presently does not want to do. When he anticipates such a change with himself, he can deploy different strategies *vis-à-vis* himself or *vis-à-vis* his future self in order to ensure that he would not in the future choose to do what his present self does not want his future self to do. Those strategies will resemble the ones he uses in an attempt to influence someone else's future behavior. In particular, he can relinquish control over certain possibilities for action and so commit his future self to behaving in one way rather than another. He may also seek someone else's assistance in attempting to gain control over his future self. "Please do not give me a cigarette when I ask for it or a desert or a second drink. Do not give me my car keys. Do not lend me money. Do not lend me a gun" are examples of pleas for this kind of help (p. 64). Schelling's general label for the kind of behavior under consideration here is "anticipatory self-command."

The discussion of self-command is related to a number of ongoing conversations in ethics such as those about *akrasia*, practical reason, self-identity, temporally extended agency, and value incommensurability, and it contains ideas which can help refresh more than one discussion. However, Schelling's account is driven by a concern not to be encountered, at least not in this exact version, in the ethics literature. The central problem for Schelling is how to accommodate the possibility of alternating preferences in a model of a rational consumer where a rational consumer is typically portrayed as someone who has uniform values and preferences over time. Human beings do not fit this model at least some of the time. The question is how to interpret cases in which the model does not work well.

There are a few contrasts that in alternative accounts are used in an attempt to help deal with analogous problems. Those contrasts include: rational versus irrational, authentic versus inauthentic, and normatively preferable versus that which must, on normative grounds, be subdued or overcome. But none of these dichotomies corresponds to Schelling's own straight versus wayward self distinction. More importantly, Schelling argues that none of them is readily applicable to human beings as we know them. Consider some illustrations.

Let us suppose that a pregnant woman chooses to deliver a baby without anesthesia, pleads for it during delivery, has it withheld, and then thanks us for that. Did we do what her real, authentic self wanted? Schelling's answer is that we do not know. He admits that if we ask such a woman who, "an hour ago was frantic with pain whether she is glad the anesthesia was denied her," she is likely to answer "yes." But he goes on, "… I don't see what that proves. If we ask her while she is in

pain, we'll get another answer" (p. 75). Both selves may be authentic, but they exist at different times. This makes intrapersonal utility comparison difficult.

Moreover, the authentic self might be the worse self, morally speaking, of the two, and there may be reasons, commonsense if not normative reasons, to suppress or help suppress in others, rather than help or let emerge, the so-called authentic self: "even if I believe that some poor inhibited creature's true self emerges only when he is drunk enough to admit he despises his wife and children and gets satisfaction out of scaring them to death, I have my own reasons for cooperating with that repressed and inhibited self that petitions me to keep him sober if I can" (p. 74).

In addition, the straight self cannot be simply equated with the rational part of us, because if we do this, it will follow that the wayward self can prevail only by dint of irrationality. However, the two competing selves can be both, in different ways, rational, just as both can be, in different ways, authentic. As Schelling says: "Suppose that trembling over the bomb's fuse made detonation extremely likely, while a calm hand could almost surely remove the fuse safely. Now should I be able to rationally persuade myself that there is no danger – none unless I create it by needlessly trembling – and not tremble?"(p. 89). If Ken creates danger by trembling, he may blame himself, indeed, for lack of bravery and for succumbing to fear. But can he plausibly blame himself for succumbing irrationally to fear?

It should be noted that Schelling does not offer a clear-cut answer to this question. Is Ken irrational if he endangers himself by trembling, or, drawing on other of Schelling's examples, by drinking salty water while at sea because he is dying of thirst though he knows full well that he thereby sentences himself to death, or by turning his back and running from a lion though he is fully convinced that the only way for him to save his life is to face the lion who, he knows, will attack him from behind? Schelling's considerations suggest that the answer is not a straightforward "yes." But that is all Schelling seeks to establish. He never asserts that the answer is a straightforward "no" either. That may seem like an evasion but in fairness, the issue is murky, and Schelling is just too careful and nuanced a writer to provide a simple response.

Finally, this account of self-command, just like the account of commitment, is meant to be descriptively correct while free from normative implications. Schelling does not discuss the question whether and why one should try to gain self-command; the problem for him is only, what happens if one does. As he puts it: "Anyone who is happily addicted to nicotine, Benzedrine, valium, chocolate, heroin, or horse racing, and anyone unhappily addicted who would not elect the pains and deprivations of withdrawal, are not my subject. I am not concerned with whether cigarettes or rich desserts are bad for you, only with the fact that there are people who wish so badly to avoid them that, if they could, they would put those commodities beyond their own reach. It is not an invariable characteristic of these activities that there is a unanimously identified good or bad behavior" (p. 69).

A plethora of questions concerning the accounts of both commitment and self-command suggest themselves. For instance, is a rational decision-maker or consumer who has competing preferences over time irrational? If not, what is rationality such that lack of preference uniformity does not entail irrationality? Is

the supposition that rationality requires uniformity of preferences over time a fiction of philosophers, of economists? If it is not a groundless fiction, what is its basis?

These are not the only grounds on which Schelling's account of commitment can be queried, surely. There are others. Consider, for instance, the intuitively plausible idea that making a commitment requires a change of heart and that a person succeeds in committing himself to something always and only when he succeeds in upholding the commitment made. Suppose Pete makes a commitment to be true to Laura. Laura does not believe him and leaves him because she does not believe him. But Pete does stay true to her. He spends his life writing poetry and music devoted to her. He stops any other interest before it takes root in his heart and dies whispering her name. There is certainly an important sense in which he succeeds in his attempt to make a commitment even if he fails strategically in Schelling's sense – the sense in which he fails to influence Laura's expectations in the desired way.

Strategies of Commitment cuts across interdisciplinary boundaries in multiple directions. That is true of Schelling's work generally. His earliest and perhaps best well-known 1960 book *The Strategy of Conflict* was, in its time, reviewed by more than fifteen academic journals from a number of areas including, besides economics, sociology, psychology, political theory, and mathematics. Interestingly enough, none of them was a philosophy journal, and the situation with Schelling's subsequent work is similar. This lack of attention on the part of philosophers is a loss for philosophy rather than for Schelling.