Explaining Norms, Geoffrey Brennan, Lina Eriksson, Robert E. Goodin and Nicholas Southwood, Oxford University Press, 2013, vii + 290 pages.

In the apt epigraph to this book Jon Elster proclaims that "[n]eoclassical economics will be dethroned if and when … sociological theory comes up with a simple and robust theory of the relation between social norms and instrumental rationality". While Geoffrey Brennan, Lina Eriksson, Robert E. Goodin and Nicholas Southwood don't quite manage to topple the throne, they are making good progress with digging up its foundations and building up a worthy challenger to the incumbent. In an unusual team effort, Brennan, Eriksson, Goodin and Southwood genuinely wrote this book together, combining their competence in economics, philosophy and political science. Even though elements of earlier, separate publications are recognizable, the book adds substantially to previous work and integrates hitherto unconnected theories.

Given its multiple authors, it is not surprising that this is not a book of one single big idea. Rather, a small number of recurrent themes and motifs connect the different chapters. *Explaining Norms* comes in three parts. Part I lays the theoretical foundations by working towards a definition of norms, and by distinguishing between formal and non-formal norms and between moral and social norms. Part II is concerned with explaining the dynamics of norms: how they emerge, persist and change. This part also addresses the debate between rational choice theorists and proponents of more interpretative, contextual approaches. Part III shows how the toolbox developed in Part I can be put to use by explaining behaviour: norm

following, conforming, breaching. BEGS also point out that norms can not only regulate observable behaviour but also attitudes, including unobserved feelings.

A detailed summary is not feasible in this short review. Instead, I will focus on a couple of interesting or controversial points. These include BEGS's definition of norms, how they distance themselves from competing theories of norms, their distinction between social and moral norms, the commitment to a mixed methodology including but not restricted to rational choice reconstructions, and their analysis of what it means to comply with and to follow a norm.

Broadly following H. L. A. Hart, BEGS offer this account of norms:

"A normative principle P is a norm within a group G if and only if:

(i) A significant proportion of the members of G have P-corresponding normative attitudes; and

(ii) A significant proportion of the members of G know that a significant proportion of the members of G have P-corresponding normative attitudes." (p. 29)

Condition (i) states BEGS's most fundamental assumption: norms are clusters of normative attitudes. These attitudes "include at least": cognitive states, normative expectations, and (dispositions to have) reactive attitudes. These attitudes are "P-corresponding" if they "reflect the *content* and *normative force* of P". In the simplest possible case, the normative attitude "just is the content of P" (all p. 29), so that condition (i) is met just in case a significant proportion of individuals holds the

judgement that P. For example, the norm that one ought to wear black at funerals would require (in the simplest possible instantiation) many members of the relevant reference group to hold the normative attitude that one ought to wear black at funerals.

Condition (ii) is a mutual knowledge constraint. It rules out situations in which everyone privately has the right sort of normative attitudes but mistakenly thinks that no one else shares these attitudes. In such cases of pluralistic ignorance, the principle P is not a norm according to BEGS's definition. Condition (ii) does not go as far as demanding *common* knowledge. The weaker mutual knowledge condition implies, counter-intuitively, that a norm can exist if, first, each group member knows that all group members have P-corresponding attitudes, but, second, each group member also believes that all other members think that no one else has these attitudes. But is it plausible to say that a norm exists in that setup? In the scenario described each group member would not expect their fellow group members to act in accordance with the norm because they are under the impression that for their fellow group members condition (ii) is violated. Given this impression, compliance is not likely, suggesting that mutual knowledge may be too weak. BEGS say that a reader for O.U.P. convinced them to drop the more demanding common knowledge assumption. I think that their original, more demanding condition would have been the right one.

BEGS provide a non-reductive account of norms. By this they mean that their definition does not replace the normative component of norms with something non-normative. They insist that the normative component is, in fact, irreducible and contrast their non-reductive account with some other reductive ones, including the

most advanced competitor theory: Cristina Bicchieri's (2006) theory of social norms. In Bicchieri's view, social norms are institutions to solve cooperation problems, or, more precisely, mixed-motive games. Simplifying Bicchieri's account a little, a behavioural rule is a social norm if there is a large enough subset of the population such that each member (1) knows about the behavioural rule; (2) prefers to comply if he believes that (2.a) enough others members of the population comply and (2.b) enough others normatively expect compliance or may sanction non-compliance. Bicchieri sharply distinguishes between the existence of a social norm and compliance with a social norm: a social norm is complied with if (2.a) and (2.b) are true. However, the norm can exist without being followed if the conditional preference (2) is in place, even though the conditions are not actually met. This account reduces social norms to beliefs and desires, but the desires are of a very special nature: they are conditional on perceived normative and empirical expectations. (Whether Bicchieri's account is truly reductive is a complicated issue. A full analysis is beyond this short review.)

To pump our intuitions against Bicchieri's proposal, BEGS provide an example: "The Chastians" (p. 25) are a chaste society, act in a chaste manner, and regard such behaviour as appropriate and expected by others. However, they are "secretly thrilled" by the idea of a much more permissive normative code. In fact, their desires are of the conditional form (2) as outlined by Bicchieri: if they believe that enough others comply with the 'lewd' code, and believe that others expect them to comply with it, then they prefer to comply with the permissive norm themselves. In Bicchieri's terms, the lewd norm exists in Chastia, though it is currently not followed. BEGS characterize this result as a reductio ad absurdum, insisting that one can, at

best, speak of a possible norm, not an existing permissive norm of sexual relations in Chastia. That seems a little fast. It is true that Bicchieri's choice of terms is unfortunate; it would have been more helpful if her 'existing-but-not-followed' norms were called "latent norms". However, apart from naming conventions, her analysis of the Chastians is spot on: Chastia has a prude norm that is followed because of widespread empirical and normative expectations, and a (latent) lewd norm that is not followed because no one expects others to comply with it or have normative expectations that the lewd norm be followed. However, Chastia could quite easily experience a normative revolution once it became apparent that most Chastians have a conditional preference in line with the lewd norm. That this analysis is available is a major advantage of Bicchieri's account. Far from being a reductio, Bicchieri's conception handles the Chastians quite well. Thus, with regard to social norms at least, I am not convinced that the BEGS definition has the edge over Bicchieri's proposal.

An important feature in *Explaining Norms* is the sharp distinction of social and moral norms, laid out in chapter 2. Put succinctly, social norms require a specific normative attitude: social norms are judgements about what one ought to do, "grounded" in perceived existing practices. Moral norms, by contrast, are never grounded in practices; they are practice-independent. The important proposal here is that it is the *justification* of the normative attitudes that distinguishes social norms: a social norm ought to be followed because others do so. It does not matter whether the practice is correctly or mistakenly perceived to be prevalent. For instance, BEGS imagine a society in which everyone strongly endorses the judgement that one ought

not to urinate in public swimming pools, everyone believes that this rule is strictly adhered to, while in fact everyone urinates in the pool. It is the perceived practice that grounds the norm, not what actually happens (though the regular publication of water quality analyses would be the downfall of the anti-urinating norm). The "grounds view" of social norms explains why many philosophers are uneasy about the influence of social norms: since social norms are grounded on perceived social practice, they may have a status quo bias, they can be parochial, backward, unjust or immoral.

Note that the "grounds" view might also be a challenge to those who think that all norms, including moral norms, are nothing but sophisticated equilibria. Since it is (arguably) pointless to comply with a norm when no one else has settled on that equilibrium, every norm appears justified only when there is an existing practice, according to these theorists. Such accounts of moral norms deny that there is any practice-independent justification and are therefore not compatible with BEGS's analysis of moral norms. Perhaps the truth in this matter is not so clear-cut: BEGS are surely right to say that the justifications of some norms (and especially social norms and conventions) are much more dependent on an existing practice than others. This is surely true for norms regulating and coordinating fashion, etiquette, and so on. These norms are justified only if the practice exists right now; they quickly loose their normative power when the practice changes. On the other side of the spectrum are norms whose justification is much less responsive to current social practices. The moral norm against killing infants would, arguably, remain justified even if killing infants was an existing social practice (though even that is controversial – think about ancient societies in which infanticide was much more common). Between these two

extremes, however, there exist norms that are clearly "grounded" in social practices, but only respond to change that is profound, long-lasting, or unlikely to be reversed.

BEGS embrace methodological pluralism in their enterprise to make sense of norms. While professing a preference for "rational reconstruction", they also identify limits to rational choice and functionalist explanations (see chs. 6 and 7). In a very helpful methodological investigation, BEGS point out that reference to the function and purpose of a norm works best when the purpose is well-defined, the norm is the most efficient way to serve the purpose, and the norm was created explicitly to serve the purpose (p. 143-4). Given these conditions, the standard concerns about functionalist explanations can be cast aside. If these conditions are not met, then game-theoretic and evolutionary accounts are of more limited applicability. They are typically still useful for explaining why a norm persists, but they do not tell us much about why and how a norm emerged. In those cases, BEGS suggest, we need to turn to the analysis of social meaning (ch. 7).

Standard rational choice models assume fixed preferences. If they take into account non-monetary or non-material desires at all, they often do so in form of generic "catch-all" placeholders like "expressive utility", "warm glow", and so on. For an understanding of norms it is useful to fill these placeholders with content, and to develop a theory of preference formation. BEGS suggest that the concepts of social roles, identity and esteem can be helpful in that regard. Norms are intimately linked with the social roles we play or the identities we adopt. In fact, many roles and identities come with a set of social norms that apply precisely to those people who play or want to play a certain social role. For example, the role of a Premier League

footballer entails following certain norms that regulate how one speaks, dresses, which attitudes one is supposed to have, and so on. Until very recently, it appears to have been part of the social role as a professional athlete not to be attracted to the same sex, or at least not to reveal in public if one is—which would explain that not a single Premier League footballer is known to be gay. If one identifies with a well-defined role it becomes hard to break out of the existing expectations, but it also requires a re-building of one's own identity, which can be a severe psychological burden. Investigating roles and identities can be the starting point for a theory of preference formation. It is often the identities we are socialized into that shape the interests and the attitudes that are taken as the input of rational choice models.

Identity-based accounts of norms are not only useful for explaining norms in the personal realm, they can also be used to analyse the diffusion of norms in international relations. Think of a state that is part of the "international human rights community" in a non-binding, non contractual sense. Identifying with the community of "civilized states" paves the way for peer pressure, leading to more and more compliance (p. 170). This is not to say that a "rogue state" would eventually be forced to fully internalize human rights norms; it means that the costs of not being identified with a certain legal community can be a subtle enforcement mechanism.

Once a norm is established in a group, what is the adequate response to it? The minimal requirement is to conform with the norm. This means that one acts in accordance with the norm's prescriptions. Conforming with the norm does not mean that one is responsive to the norm (p. 195), the conformity might be coincidental or perhaps based on purely instrumental reasoning such as the avoidance of punishment.

In light of this, BEGS suggest a notion of responsiveness to a norm that they call "norm following". Roughly, an individual follows a norm if the norm itself provides the relevant reason for conformity. But how can this view be fleshed out more precisely? One option is to understand this in causal fashion: following a norm means desiring to comply with the norm because the norm requires this action. This suggests that the norm is the difference-making cause for the compliance desire. In all nearby possible worlds in which the norm exists the agent conforms, and in all nearby possible worlds in which the norm does not exist the agent does not. BEGS reject this proposal because they think that one can internalize and follow a norm in a rigid way, conforming even in those nearby possible worlds in which the norm does not exist (p. 200). That is plausible: it would be odd to say that an individual does not truly follow a norm if she is not perfectly responsive to the (non-)existence of the norm, as the internalization of norms often goes hand-in-hand with automated scripts.

In light of this, BEGS suggest a different definition of norm following: one follows a norm if one complies with the norm, "has a non-instrumental desire to act in accordance" (p. 202) with that norm, and this non-instrumental desire *explains* one's conforming. BEGS eschew "causal talk" and prefer "explanatory talk" in their definition of norm following. However, it seems to me that a small tweak to the causal story would work at least equally well. The proposal is: following a norm means conforming with the norm and desiring to conform with it *because one believes that the norm requires this action in the situation*. One major advantage of my causally geared definition is the explicit consideration of the positive and negative tracking component of the causal relation. The causal relation demands that an individual acts in accordance with the norm in the nearby possible worlds in which

the individual believes that this is prescribed by the norm, and does not act in accordance with the norm in the nearby possible worlds in which that belief is not present. This is plausible – an individual failing to be responsive to a norm in that sense can hardly be said to follow a norm. The refined causal notion of norm following is also immune to the earlier objection that causal notions cannot account for rigid norm following. Since the cause is now the belief that the norm exists (rather than the existence of the norm itself), rigid norm following is the result of incorrect beliefs about the applicability of a norm; a plausible interpretation.

For an example of a violation of positive tracking, consider lazy Larry. Larry has a non-instrumental desire to comply with the norm to wear black at funerals and this explains why he puts on his black trousers rather than his jeans before he heads to the funeral. However, suppose we also know that if Larry's only black trousers were at the dry cleaner he would not make the faintest effort to find an alternative pair and instead wear his blue jeans. Moreover, he would be totally unconcerned about this. It seems that Larry follows a norm in the BEGS sense (though this depends on how they spell out the notion of "explaining"), while my causal account provides resources to maintain – correctly, I think – that Larry's desire to conform is not sufficiently robust to say that he follows the norm.

The points discussed in this short review provide only a glimpse of the philosophical and empirical richness of *Explaining Norms*; it is by no means a summary of all the arguments offered. Even where I disagree with BEGS, I hope to have shown that I found their arguments enlightening and challenging. This is a book packed with ideas and a treasure trove of empirical examples and anecdotes. It will

surely propel the debate about norms in the social sciences and philosophy. One particular strength of the book is the tight interaction between theory and practice. Bookseller and librarians will have a hard time deciding whether this book goes into the "Economics", "Politics", "Sociology" or "Philosophy" section. It probably belongs in all four. In fact, *Explaining Norms* should be read widely by scholars and students of those and other disciplines.

Kai Spiekermann[*]

REFERENCES

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

BIBLIOGRAPHIC INFORMATION

Kai Spiekermann is Associate Professor of Political Philosophy at the London School of Economics. Among his research interests are normative and positive political theory, philosophy of the social sciences, social epistemology and environmental change. His recent publications have focused on mechanisms of norm avoidance and moral knowledge, jury theorems, and on reductionism and holism in the social sciences.

---

[*]London School of Economics, Dept. of Government, Houghton Street, London WC2A 2AE, UK. Email: k.spiekermann@lse.ac.uk. URL: http://personal.lse.ac.uk/spiekerk