interpretation of social behaviour is the foundational bedrock of all social scientific explanation. Merely pointing out the causes of economic phenomena will not do unless the causes themselves are understandable through our emphatic identification with the beliefs and motivations of the actors involved. But such intentional fundamentalism is incompatible with the idea that the social sciences should be about finding the causes and mechanisms behind social phenomena, regardless of whether or not these causes and mechanisms are described in intentional, sub-personal or macro vocabulary. The fundamentalist stance severs economics from the other sciences, such as psychology, in having a totally different standard of explaining things. If one carries Lagueux's argumentation to its logical conclusion, economics is and should remain a separate science, but one more akin to the humanities than to the rest of the sciences.

Jaakko Kuorikoski
University of Helsinki

REFERENCES

Henderson, D. 2010. Explanation and rationality naturalized. *Philosophy of the Social Sciences* 40: 30–58.
Kitcher, P. 1993. *The Advancement of Science*. Oxford: Oxford University Press.
Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.
Vromen, J. 1995. *Economic Evolution*. Abingdon: Routledge.

doi:10.1017/S0266267111000113

Fairness, Responsibility, and Welfare, Marc Fleurbaey. Oxford University Press, 2008. x + 295 pages.

Responsibility-sensitive egalitarianism, especially of the luck-egalitarian kind, has played a prominent role in recent discussions of distributive justice. The core idea is that there are some factors for which people are responsible – say, their productive efforts – and some factors for which they are not – say, their talents. (Some responsibility-sensitive egalitarians withhold judgement on whether, ultimately, anyone is ever responsible for anything, but their writings are based on the working assumption that some individuals are sometimes responsible for being worse off.) Unlike outcome egalitarians, responsibility-sensitive egalitarians allow that inequalities reflecting differences in the former factors may be just. Like outcome egalitarians, responsibility-sensitive egalitarians contend that inequalities reflecting differences in the latter are unjust. Luck-egalitarians see it as a strength of their theory that it accommodates widespread, seemingly anti-egalitarian intuitions to the

effect that justice does not require, indeed in some versions, forbids, (involuntary) compensation of the lazy worse off by the hard-working better off, while affirming the egalitarian credo that it is unfair for some to be worse off than others for reasons over which they had no control. Other responsibility-sensitive egalitarians are less enthusiastic about this accommodation.

Marc Fleurbaey's book, *Fairness, Responsibility, and Welfare* (henceforth: FRW) is a major contribution to the literature on responsibility-sensitive egalitarianism, and on distributive justice in general. Over the years, Fleurbaey has contributed significantly to these areas. He is well known for his harsh objection to standard forms of luck-egalitarianism. This objection points out that these positions imply that people who end up in dire straits through their own choice – e.g. unlucky Bart who, for once, makes the risky choice of not wearing a helmet while driving a motorcycle and has a traffic accident leaving him in urgent need of medical assistance (153) – have no claim to compensation. While FRW brings together much of the material developed in Fleurbaey's previously published articles, FRW is a free-standing piece of work. It offers an in-depth account of Fleurbaey's responsibility-sensitive egalitarianism and is distinctive for the author's impressive combination of mathematically informed analytical skills and nuanced egalitarian intuitions.

With a few exceptions, contributors to the literature on responsibility-sensitive egalitarianism come from either economics or philosophy. FRW aims to bring together these two disciplines. This ambition is inspired by the 'conviction that such a dialogue can produce results that cannot be achieved when each discipline works in isolation' (2). One practical obstacle to such dialogue is mathematics, which goes down less well with philosophers than with economists. Accordingly, FRW is separated into formal and non-formal sections, the former being optional and the latter being largely free from formalism. While this makes (large parts of) FRW readable to philosophers, still, they will not see it as an easy read. In part, this reflects that FRW is not written *for* an audience that (i) is acquainted with the philosophical literature on equality, and that (ii), with this acquaintance as a point of departure, wants to see its main claims analysed and assessed with the economist's tools suitably simplified for presentational purposes.

FRW has two main purposes: 'to develop a theory of the distributive implications of holding individuals partly responsible for their situation' (2–3) and to suggest which forms of responsibility-sensitive egalitarianism are the most attractive (3). For these purposes, it is essential to explore two logically independent and potentially conflicting principles: 'the compensation principle saying that inequalities not due to responsibility should be eliminated and the [liberal] 'reward' principle saying that inequalities due to responsibility should be left untouched' (7). Since both

of these principles refer to inequalities (not) due to responsibility, their implications depend crucially on what these are. Normally, people do not determine the level of the relevant *equalisandum* (which, for simplicity, I assume is wellbeing) of others, but might to some extent determine their own level. Hence, one might suggest that what matters is not whether individuals are responsible for the inequalities between them, but whether each is responsible for her own level of wellbeing. Whether this is so may hinge on whether the individual's level of wellbeing is caused by her own choices. But this needs to be spelled out more carefully. First, one might be responsible for a low level of wellbeing because of non-choice involving negligence. Second, as Fleurbaey points out, choices do not simply produce outcomes in isolation from the context of choice, and which context justice requires is an issue to which the idea of responsibility does not speak.

Fleurbaey thinks that while the idea that people ought to be held responsible, e.g. for their choices, is plausible, it tends to be misused to justify unfair inequalities. Specifically, he distinguishes between two different reward principles for the apportionment of personal outcomes to personal responsibility (10): the mentioned liberal reward principle and a utilitarian reward principle. The former principle, which many responsibility-sensitive egalitarians endorse unreflectively, says that once inequalities due to differences in non-responsibility related factors have been eliminated, no further redistribution is required. The latter principle recommends redistribution beyond this point, namely redistributions that 'enhance the total outcome of individuals similarly situated with respect to circumstances' (10). Compare the following two simple distributions: one in which two people with high efforts have 12 each and two people with low efforts have 0 each, and one in which two people with high efforts have 8 and two people with low efforts 7. The liberal reward principle is indifferent between these two distributions, whereas the utilitarian reward principle prefers the latter to the former. If there is one single, most important message of FRW, it is that responsibility-sensitive egalitarians should pay greater attention to the issue of the correct reward principle and that the liberal one cannot be taken for granted.

FRW starts with an explanation of what fairness amounts to in a simple analytical context. It nicely demonstrates the point just made showing why the principle of compensating for non-responsibility factors underdetermines which distribution should be selected. A distribution where all individuals with low efforts end up with the same outcome whatever their non-responsibility characteristics and all individuals with high efforts end up at the same, and *lower* level suppresses inequalities for which individuals are not responsible. It seems natural to neutralize across all individuals the effects of luck by equalizing factors for which they are not responsible, but neutralizing the effects of luck across individuals

with the same responsibility characteristics also satisfies the compensation principle.

This crucial point relates to Susan Hurley's critique of the luck-neutralizing fallacy. Hurley (2003: 146–180) thinks that luck-egalitarians fallaciously infer from the fact that the effects of luck, i.e. factors for which people are not responsible, have been neutralized that the resulting distribution must be an equal one. According to Hurley, this does not follow, because the baseline from which luck-induced deviations are neutralized may not be an equal one, but one that, say, maximizes the sum of wellbeing. In response, Fleurbaey points out that the principle of Equal Wellbeing for Equal Responsibility – 'Two individuals with identical responsibility characteristics should have the same level of wellbeing' (25) – is logically equivalent to the principle of Circumstance Neutralization – 'At the selected allocation it should be possible to express individual wellbeing as a function of responsibility characteristics only' (26) – and the latter clearly rules out the sort of scenario that Hurley employs in her critique. If the baseline situation is not an equal one but, say, an unequal one where individuals who are better at transforming resources into wellbeing have more resources than others, then individuals' wellbeing cannot be expressed as a function of responsibility characteristics *only*. Accordingly, luck-egalitarians can appeal to the principle of Equal Wellbeing for Equal Responsibility in response to Hurley.

Chapter 2 builds on the model introduced in the previous chapter and shows why the principle of Equal Wellbeing for Equal Responsibility might be incompatible with the principle of Equal Treatment for Equal Circumstances – 'Two individuals with identical circumstances should be submitted to the same transfer' (31) – despite the fact that both derive from the no-envy test for equality (43). Incompatibility arises in situations where low non-responsibility factors reduce the positive impact of responsibility factors. Here, achieving no-envy between individuals is generally impossible. Fleurbaey critically explores various responses, including the strategy of weakening either principle or of satisfying the no-envy test as much as possible, e.g. by minimizing the number of envy occurrences etc.

The ensuing chapters 3 to 5 introduce a number of complications to the initial analytical model. Chapter 3 discusses incentives constraints which render it impossible to implement allocation rules perfectly and, accordingly, considers which second-best rules should be adopted. For instance, a rule equalizing income would face the constraint that by rendering income independent of effort, incentives to work are weakened and as a result everyone may end up worse off than they would have been under a laissez-faire scenario involving no redistribution. While Fleurbaey's concerns in this chapter are not directly related to Cohen's critique of the incentive justification for inequality, the latter has had

a central role in recent political philosophy, and from this perspective, it would have been nice to see what Fleurbaey has to say about the issues this critique raises. Presumably, what would qualify as perfect allocations on Fleurbaey's account – because planners can costlessly observe individual characteristics etc. – may not qualify as just allocations on Cohen's account – because (talented) individuals ought to make efforts even in the absence of their having incentives.

Chapter 4 investigates which distributive implications unequal skills should have, provided individuals are not responsible for their earning capacity, but fully responsible 'for their preferences over consumption and leisure' (101). It shows that, under some circumstances, finding an envy-free and efficient allocation of working hours is impossible. This impossibility is tied 'to an internal conflict between no-envy among individuals with identical circumstances and no-envy among individuals with identical responsibility' (107). Fleurbaey offers several solutions that come as close as possible to satisfying the relevant desiderata, including avoiding the slavery of talented.

Chapter 5 explores income redistribution under the assumptions that fairness requires redistribution from high-skilled to low-skilled individuals, and that only gross and net income is observable. The discussion of Van Parijs' claim – that real freedom for all 'implies focusing on the payment of a large basic income, and ignoring what happens to disposable income for positive levels of earnings' (151) – is particularly interesting. The chapter also observes that jobs occupied by people with higher skills tend to be more pleasant and to carry various symbolic advantages. In Fleurbaey's view, this implies that 'it is questionable to hold individuals responsible for their preferences over leisure and consumption. Those who only have access to unpleasant jobs will naturally be more averse to work and should not be held responsible for this' (148). I did not quite see the reasoning here. The fact adduced by Fleurbaey does not suggest that people with high-skill jobs should not be held responsible for their preferences over leisure and consumption compared to other people with high-skilled jobs. Nor does it suggest that one should not hold those (few?) people with high-skill jobs that choose to work for fewer hours than people with unpleasant jobs responsible for their relevant choices. Third, it might be that people with high-skilled jobs tend to have more pleasant leisure – fancy holidays, gourmet restaurants or a beautiful mansion – in a way that counterbalances the greater attractions of their jobs. Finally, in one sense of responsibility, the mere fact that one has a job that is less pleasant than that of others does not mean that one is not responsible for the number of hours one works. That choice is attributable to oneself, and one can be asked to account for one's choice regardless of what others do. So in this non-comparative sense, one *is* responsible for one's choices regarding leisure and consumption,

even if the issue of which harms and benefits should accrue to one as a result of that choice is a different matter. (The distinction employed here is Scanlon's distinction between attributive and substantive responsibility.)

Chapter 6 addresses the issue of the fair allocation in cases where individuals have some degree of control over the amount and kind of risk to which they expose themselves. Of great interest here is Fleurbaey's reconstruction of Dworkin's influential distinction between brute and option luck: 'Option luck is a matter of how deliberate and calculated gambles turn out—whether someone gains or loses through accepting an isolated risk he or she should have anticipated and might have declined'. Brute luck is 'a matter of how risks fall out that are not in that sense deliberate gambles' (Dworkin 2000: 73). If I suddenly go blind as a result of a genetic condition, my brute luck is bad, but if I buy a lottery ticket and win, my option luck is good. According to Dworkin (roughly), justice requires that bad brute luck is compensated and that outcomes reflecting differential option luck are allowed to persist. Fleurbaey denies the latter claim and defends the principle that individuals should never be held responsible for their luck, good or bad (154). While some may prefer the risk involved in option luck, no one prefers gambling *and* losing. Moreover, no one controls whether they have bad and good option luck. Dworkin will respond that if the results of differential option luck are eliminated, then, in effect, those who prefer a risky life style have been prevented from living the sort of life they prefer. Fleurbaey concedes this point, but convincingly retorts that its significance should be seen from an *ex post* perspective. That is, from this perspective, not only *ex post* gains matter: '*ex ante* satisfaction' of risky activities and 'basic freedoms such as the freedom to organize games matter' (158). The chapter ends with an *ex post* perspective-based critique of Dworkin's insurance device for modelling an endowment-insensitive distribution.

Chapter 7 picks up the thread of the harshness objection. What does distributive justice require in the case of individuals who regret their past reckless choices and want a 'fresh start'? The harsh view says that because they are responsible for their past choices, they have no claim of justice for assistance even if, say, the virtues of benevolence or compassion favour helping them out (perhaps even if they do not regret their past choices, and perhaps even if they do not regret the regrettable consequences of their past reckless decisions). Fleurbaey defends the softer view that justice is compatible with the sharing among the whole community the costs of reckless choices (178). (It is not clear to me if, in FRW, Fleurbaey accepts the stronger claim that justice requires such cost sharing.) A fresh start policy imposes 'constraints on initial decisions' (195), but there is, Fleurbaey argues forcefully, no strong ethical intuition to the effect that fresh starts are questionable. A choice of rules allowing this is clearly not a choice of a less fair set of rules than the choice of a set of rules

that do not allow fresh starts (180). While this is correct, I suspect the underlying intuition here is that, within a very broad range of rules, any set of rules are compatible with justice, provided they are announced in advance and participants consent. No such condition may be satisfied in real-life distribution scenarios.

Chapter 8 develops the contrast between the liberal and utilitarian principles of compensation and explores various versions of the former, while chapter 9 narrows the perspective of social evaluation of distributions from a global to a local perspective, where one is concerned with a partial aspect of a distribution, e.g. intergenerational social mobility. Chapter 10 contains Fleurbaey's contribution to the 'equality of what'-debate, that is, the debate about what it is that people should be equal with regard to. Fleurbaey characterizes his position as a theory of equal autonomy. An important difference between this position and luck-egalitarian theories is that it gives a less prominent role to responsibility (273). The difference comes out in relation to the 'serial squanderer' where luck-egalitarians appealing to a pre-institutional notion of responsibility denies the case for compensation on grounds of justice, whereas Fleurbaey's freedom-based approach involves a claim of justice to 'equal status' and 'a basic bundle of freedoms' (275). He also responds to the egalitarians who have pressed against luck-egalitarians the point that what egalitarians should be concerned with is not distribution (of material goods), but the non-hierarchical character of social relations. Fleurbaey shares this concern – in fact, he aptly ties it to his endorsement of fresh start policies – but submits that a concern for distribution includes a concern for 'goods that go with social relations' (245).

The chapter also scrutinizes the cut between factors for which individuals are responsible and those for which they are not, a distinction that has simply been taken as given in previous chapters. Fleurbaey is reluctant to ground the cut on metaphysical issues of free will that are unlikely to be resolved anytime soon, thereby leading to the impotence of a theory of distributive justice. This is not to deny the truth of Cohen's claim that justice *is grounded* on metaphysical issues of free will, but Fleurbaey canvasses some additional concerns along the lines of those embodied in the Rawlsian idea of public reason that more clearly contradict it.

Finally, Fleurbaey helpfully contrasts preference- and control-based conceptions of free will and responsibility, suggesting that control-based accounts conflict with economic models of choice since they see 'individual decisions as a mechanical optimization' (251). I am uncertain why this is the case. On some compatibilist, control-based accounts of free will (and moral responsibility), free will is a matter of acting from a reasons-responsive mechanism, and provided (the

admittedly controversial assumption) that one's preferences provide reasons, mechanical optimization may proceed from a reasons-responsive mechanism.

All in all, FRW is an excellent contribution to the field of distributive justice well worth the considerable efforts I suspect reading it requires in cases like my own: that of a philosopher with a regrettably rudimentary grasp of economics.

Kasper Lippert-Rasmussen
University of Århus

REFERENCES

Dworkin, R. 2000. *Sovereign Virtue*. Cambridge, MA: Harvard University Press.
Hurley, S. 2003. *Justice, Luck and Knowledge*. Cambridge, MA: Harvard University Press.

https://doi.org/10.1017/S0266267111000113 Published online by Cambridge University Press