



The opportunity criterion

Robert Sugden

Paper prepared for NIBS 2016 conference 'Assessing well-being when preferences are incoherent', University of East Anglia, 4–6 April 2016

The history of this project:

From the mid-1970s I have written about the descriptive and normative limitations of neoclassical rational choice theory.

I have argued that there are important human motivations that that theory is not able to represent, including:

- reciprocity/ team reasoning;
- the idea that liberty/ opportunity is valuable;
- the influence of psychological factors on choice [work with Graham Loomes].

I've never thought of these motivations as incompatible with one another – in fact I have come to see David Hume's *Treatise of Human Nature* (1740) as a founding text for the study of all three.

For the last 20 years, I've worked on the problem of finding a form of normative economics that is compatible with the findings of behavioural economics.

My approach uses concepts of reciprocity and opportunity. I see it as in the classical liberal tradition of Hume, Smith and Mill (all of whom wrote before the invention of rational choice theory).

I've been arguing for this approach since 2004...

[E.g. *American Economic Review* 2004; *Social Choice and Welfare* 2007; *Social Choice and Welfare* 2012 (with Ben McQuillin)]

... but I don't seem to have gained many converts.

I'm now completing a book which brings the whole argument together.

This talk gives an overview of the argument.

The problem of reconciling normative and behavioural economics

From the 1940s to the 1990s, there was a consensus about how to do normative economics: **neoclassical welfare economics**.

Neoclassical welfare economics uses preference satisfaction as its normative criterion.

Presupposition: individuals have **integrated** preferences (i.e. well-defined, stable, context-independent) which are revealed in their choices.

But findings of behavioural economics show that individuals' choices often *don't* reveal integrated preferences.

Instead: predictable context-dependent effects that can be explained by psychological theories but seem not to correspond with good reasons for differences in behaviour (e.g. Kahneman, Knetsch and Thaler, 1990: median valuation of coffee mug is \$7.12 when the problem is framed as *selling*, \$3.12 when exactly the same problem is framed as *choosing*).

How should normative economics deal with this problem?

Emerging new consensus among behavioural economists: **behavioural welfare economics.**

[Proposed by:

Bleichrodt, Pinto-Prades and Wakker (2001)

Sunstein and Thaler (2003, 2008)

Camerer, Issacharoff, Loewenstein, O'Donoghue and Rabin (2003)

Bernheim and Rangel (2009)

Kőszegi and Rabin (2007);

Salant and Rubinstein (2008).

Characterisation and critique of this approach: Infante, Lecouteux and Sugden, *Journal of Economic Methodology* 2016.]

In behavioural welfare economics:

Normative criterion = individuals' subjective judgements about their own well-being. (Aim is to '**make choosers better off, as judged by themselves**' – Thaler and Sunstein, 2008.) Note: as in standard interpretation of neoclassical welfare economics.

But: it is *not* assumed that these judgements are revealed in choice.

Problem: how to define and measure individuals' subjective judgements about well-being so that these judgements, unlike preferences, are integrated. (If they were context-dependent, we would be back at square one.) Supposed solution is 'purification' of revealed preferences to eliminate 'error' and so uncover **latent preferences**.

Gerardo Infante, Guilhem Lecouteux and I argue that this 'solution' is fundamentally misconceived. It implicitly assumes the model of an **inner rational agent** with neoclassical preferences, interacting with the world through an error-prone psychological shell. But this model has no psychological foundations – we have no grounds for assuming the existence of integrated latent preferences.

My proposed approach: (1) the addressee

First question for any proposed form of normative economics:

Who is the addressee?

Normative economics produces recommendations. But who are the recommendations addressed to? And what is he/she/they supposed to do with them?

Implicit answer in neoclassical *and* behavioural economics: the **social planner**, aka the policy-maker, the legislator, the choice architect.

To explain why the social planner finds the recommendations useful, we need the implicit assumptions that he/she is:

- benevolent (hence, concerned about welfare);
- unconstrained (hence, able to act on recommendations);
- receptive to the welfare economist's advice (i.e. endorses the normative criteria and empirical analysis used by the economist).

In effect: the assumption is that the economist is the favoured adviser of a benevolent despot.

This is not an assumption about how the world really is (or even about how it ought to be). It is a **convention of economic writing**.

Since it is such a peculiar convention, economists ought to be open to the possibility of using different conventions.

The first step in my approach is to propose an alternative convention (used by James Buchanan (and, I claim, by Hobbes and Hume). This is **contractarianism**.

Contractarian recommendations are addressed to **citizens together, as potential parties to mutually beneficial agreements.**

Compare a recommendation to a social planner:

- R is in the interests of the society for which you are making decisions.

with a contractarian recommendation to citizens:

- It is in the interests of each of you separately that all of you agree to implement R.

Crucial difference in relation to paternalism...

E.g. alcohol consumption. A health expert can present evidence to a social planner and say:

- Having looked at this evidence, you can see that many people drink too much; they don't see this; so I recommend you to restrict their choices in their own interests.

But if the expert addresses citizens, she has to show them the evidence and say:

- Having looked at this evidence, *you* can see that *you* drink too much; so I recommend *you* to drink less.

(This is a perfectly sensible kind of non-paternalistic recommendation – it's odd that economists always want to address imaginary social planners.)

Or:

- Having looked at this evidence, *you* can now see that *you* drink too much; but you know you are unable to act in your own interests; so I recommend you to ask the government to restrict your choices in your own interests.

Implication: a contractarian can recommend paternalism only when individuals have self-acknowledged self-control problems.

My conjecture is that self-acknowledged self-control problems are a tiny fraction of the cases of context-dependent preferences that normative economics has to deal with.

[Behavioural welfare economists tend to see self-acknowledged self-control problems everywhere – e.g. Fudenberg and Levine's (*AER* 2006) explanation of high level of risk aversion in lab experiments: student subjects deliberately constrain their access to cash to solve self-control problems, then use experiments as a way of evading their own constraints.

But this preoccupation with self-control may be a side-effect of using the inner rational agent model. If you use that model, you have to explain why people predictably act contrary to their latent preferences, and lack of self-control is a candidate explanation.]

How common are self-acknowledged self-control problems?

News item (*The Independent* 24/03/2016)

Bookie loses £2m as panicked online gamblers lock their own accounts

William Hill warned its profits are being damaged by a new feature on its website

... As part of a new feature introduced to [William Hill's online betting website], punters can choose to lock themselves out of their accounts for 24 hours, a week, a month or up to six weeks with just the click of a button. The option was introduced by the UK Gambling Commission last November as part of a raft of measures designed to improve the industry's social responsibility. And yesterday William Hill's chief executive James Henderson said the number opting to take a break had risen by 50 per cent since the turn of the year, with about 3,000 accounts now being blocked every week. ... William Hill's online business, with 2.7 million users, made £126.5m last year.

My proposed approach (2): the normative criterion

If you (the economist) are addressing social planners, you need an operational criterion that *social planners* might plausibly want to use.

But if you are addressing citizens, you need an operational criterion that *citizens* might plausibly want to use in private choice, and/or that *citizens* might plausibly want policy-makers to use when acting as their agents.

I claim: ‘**more opportunity is better than less**’ is a plausible contractarian criterion. As a citizen, you can want to use this criterion (and want it to be used by policy-makers acting as your agents) even if you expect your preferences to be context-dependent.

Unless you have a self-acknowledged self-control problem, why *wouldn't* you want to have more opportunities to choose to have what you want (when the time comes to choose)?

However, 'more opportunity is better than less' *for individuals separately* doesn't take us very far. The problem is that in a market economy, each individual's opportunity set depends on the choices that other individuals make from theirs.

E.g. my opportunity set depends on the wage rate for the type of labour that I can supply, but that depends on whether other individuals choose to buy the goods and services it can produce.

A contractarian response to this problem: think of opportunities for individuals jointly to get what they want.

Hence: the (interactive) **opportunity criterion**.

This is a contractarian analogue of the Pareto principle, but applies to individuals' opportunity sets and makes no reference to preferences.

Opportunity criterion is satisfied iff: (i) individuals' actual choices from their opportunity sets are jointly feasible, and (ii) for every feasible transaction T among any given set S of individuals: *either* each member of S is free to choose his component of T *or* there is some member of S whose component of T is unambiguously worse than some other option he is free to choose.

I.e. If T is a transaction that all parties to it could plausibly want to participate in, each party is free to choose his component of it.

Or (with 'willing to pay for x ' = 'willing to give up what would induce others to supply x '):

Each individual is able to get whatever he wants and is willing to pay for.

(To be read as: Whatever his preferences happen to be, he is able to get what he wants. Not: Given what his preferences are, it happens to be the case that he is able to get what he wants.)

My proposed approach (3): the normative status of competitive markets

Analogue of the first fundamental theorem of welfare economics:

In any exchange economy, if for every good there is a single price ('law of one price') and if all markets clear ('law of supply and demand'), the opportunity criterion is satisfied (Sugden, *AER* 2004).

Notice that this doesn't require that consumers act on integrated preferences, only that they are price-sensitive, and that there are arbitrageurs who seek profit and don't collude.

The proof is simple. (I don't claim to be producing new, mathematically sophisticated theorems. I am showing how existing theory can be reformulated to exclude assumptions about integrated preferences.)

Proof: Consider an exchange economy in which (i) all markets clear and (ii) for every non-money good there is a single money price.

Because of (i), individuals' actual choices are jointly feasible.

Consider any feasible transaction T for any set S of individuals. Because T is feasible for S , total net acquisition of each good, summed over all members of S , is zero. So, valued at market prices, net acquisitions by members of S sum to zero.

Thus, either Case 1 or Case 2 must hold:

Case 1: for each member of S separately, the market value of net acquisitions is zero. So each person is free to choose his component of T by trading at market prices.

Case 2: there is some member of S for whom the market value of net acquisitions is negative. So his component of T is unambiguously worse than something else he could choose by trading at market prices.

[Opportunity criterion is satisfied iff: (i) individuals' actual choices from their opportunity sets are jointly feasible, and (ii) for every feasible transaction T among any given set S of individuals: *either* each member of S is free to choose his component of T *or* there is some member of S whose component of T is unambiguously worse than some other option he is free to choose.]

The result can be extended to a multi-period model: if individuals' preferences are inconsistent between time periods, the theorem still applies (McQuillin and Sugden, *Social Choice and Welfare*, 2012):

Each individual is able to get whatever he wants and is willing to pay for, *when he wants it and is willing to pay for it.*

My proposed approach (4): consumers' surplus

Consumers' surplus is a key concept in neoclassical welfare economics. Standard definition of consumers' surplus (= compensating variation) is in terms of integrated preferences.

Can we redefine the concept so that it doesn't presuppose integrated preferences?

I suggest we build on the approach used by Jules Dupuit (1844), the inventor of consumers' surplus and of cost-benefit analysis.

Dupuit has a very behavioural theory of consumer choice. E.g. the demand function 'is not known for any commodity, and it can even be said that it will never be known since it depends on the volatile will of human beings; it is today no longer what it was yesterday'. Price discrimination works by setting 'traps for the buyer's vanity and his credulity'.

Dupuit's slogan: 'Hence the saying which we shall often repeat because it is often forgotten: the only real utility is that which people are willing to pay for.'

I think this means: for the purposes of economics, there is nothing more fundamental than willingness to pay.

Dupuit's fundamental concept of consumers' surplus is the maximum revenue that can be appropriated by discriminatory prices (less what consumers actually pay). This doesn't assume integrated preferences (e.g. this measure can aggregate over what an individual is willing to pay in different contexts).

Dupuit's 'rational principles' for the building and pricing of public works:

- Build public works if willingness to pay (measured by maximum yield of discriminatory pricing) exceeds cost;
- Finance public works by discriminatory pricing.

Notice that these are principles which intermeditate mutually beneficial transactions, i.e. give people what they want and are willing to pay for, when they want it and are willing to pay for it.

My proposed approach (5): the ethic of mutual advantage

In my model of reciprocity/ team reasoning, individuals may be motivated to participate in mutually beneficial practices, conditional on others being willing too.

[‘Reciprocity’, *Economic Journal* 1984; ‘Thinking as a team: towards an explanation of non-selfish behaviour’, *Social Philosophy and Policy* 1993; ‘Team reasoning and intentional cooperation for mutual benefit’, *Journal of Social Ontology* 2015.]

This motivation can induce non-self-interested behaviour (e.g. returning trust, contributing to public goods).

But the motivation is *not* ‘kindness’ (i.e. wanting to benefit others), or ‘reciprocity’ in the social-preference sense (i.e. wanting to be kind to people who are kind to you).

It is: Playing your part in practices that (when followed by others too) benefit you.

Consider an everyday market transaction: I want my house painted; I agree a price with a decorator; she paints my house; she sends me the bill; I pay.

What is her intention in painting the house? What is my intention in paying?

She is not being kind to me: she expects to be paid, and if she is paid, she will benefit overall.

I am not being kind to her: I realise that if I try to avoid paying, I will probably lose overall.

But our intentions need not be self-interested. Each of us is playing his/her part in a mutually beneficial practice, on the expectation that the other will too. **That could be what we intend.**

Implication: Ordinary market behaviour (as typically represented in economics) can express the same kinds of intentions and motivations as other forms of cooperation (e.g. returning trust; voluntary contributions to public goods).

[Bruni and Sugden, 'Fraternity: why the market need not be a morally free zone', *Economics and Philosophy* 2008; Bruni and Sugden, 'Reclaiming virtue ethics for economics', *Journal of Economic Perspectives* 2013)]

[Contrary to a lot of the rhetoric of behavioural economics. E.g. Kahneman, Knetsch and Thaler (1986), reporting their dictator game results as evidence of preferences for fairness, contrast this with the position of 'true-believer' economists:

'... a resistance to explanations of economic actions in moral terms that has deep roots in the history of the discipline. The central insight that gave rise to modern economics is that the common good is well served by the free actions of self-interested agents in a market.']

I have outlined various elements of a proposed approach to the problem of reconciling normative and behavioural economics.

A common feature: **economic life is treated as mutually beneficial cooperation between individuals.**

Hence the provisional title of my book (which may not be approved by publishers):

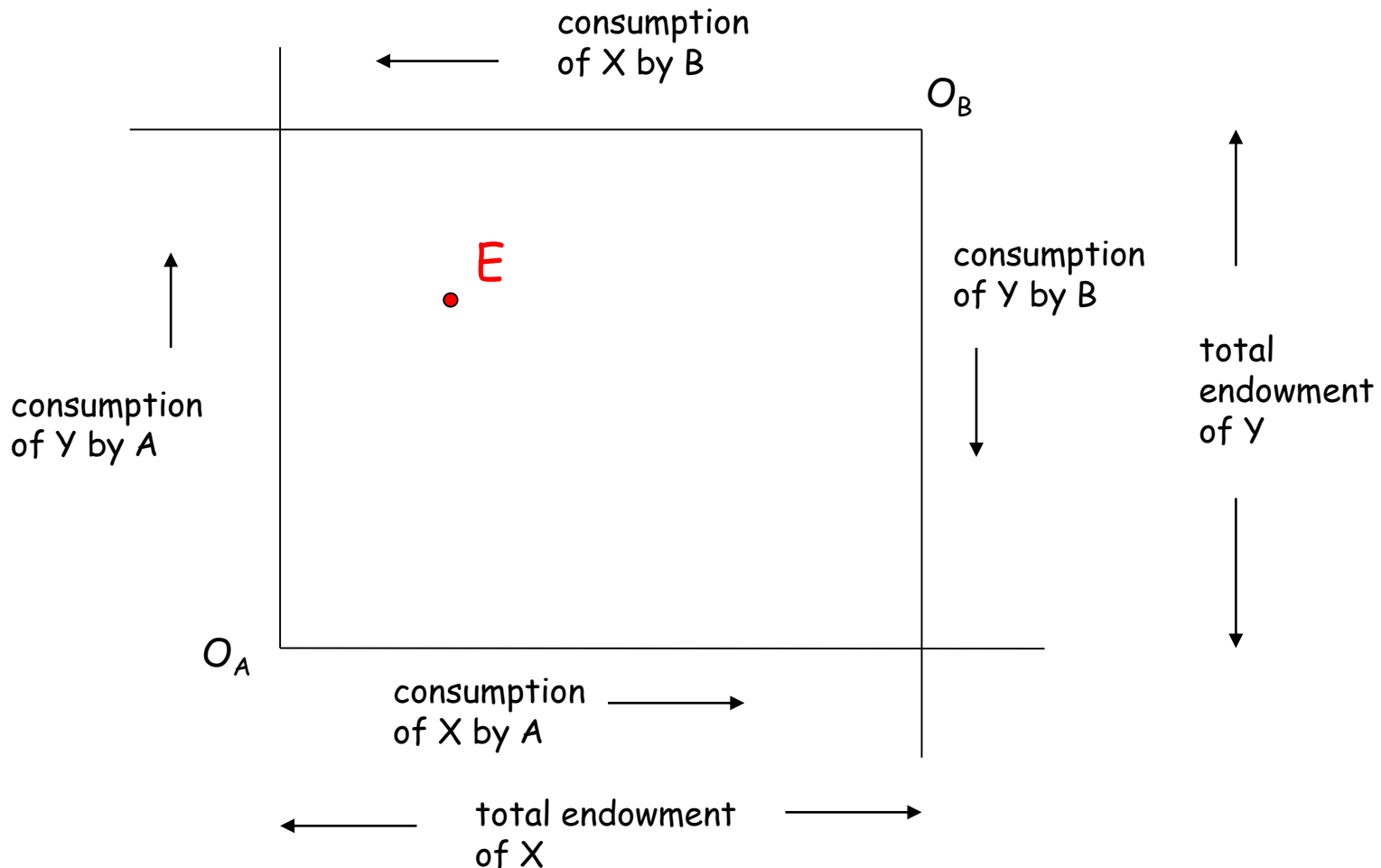
The community of advantage: understanding economic cooperation without assuming rational actors.

I hope it will be published soon, and that you will want to read it.

Thank you for listening.

Sketch of proof of theorem for simple economy

Edgeworth box: dimensions are total endowments; one point in box (E) represents individuals' endowments; any point in the box is a feasible consumption outcome. Y = 'money', which everyone is assumed to want.



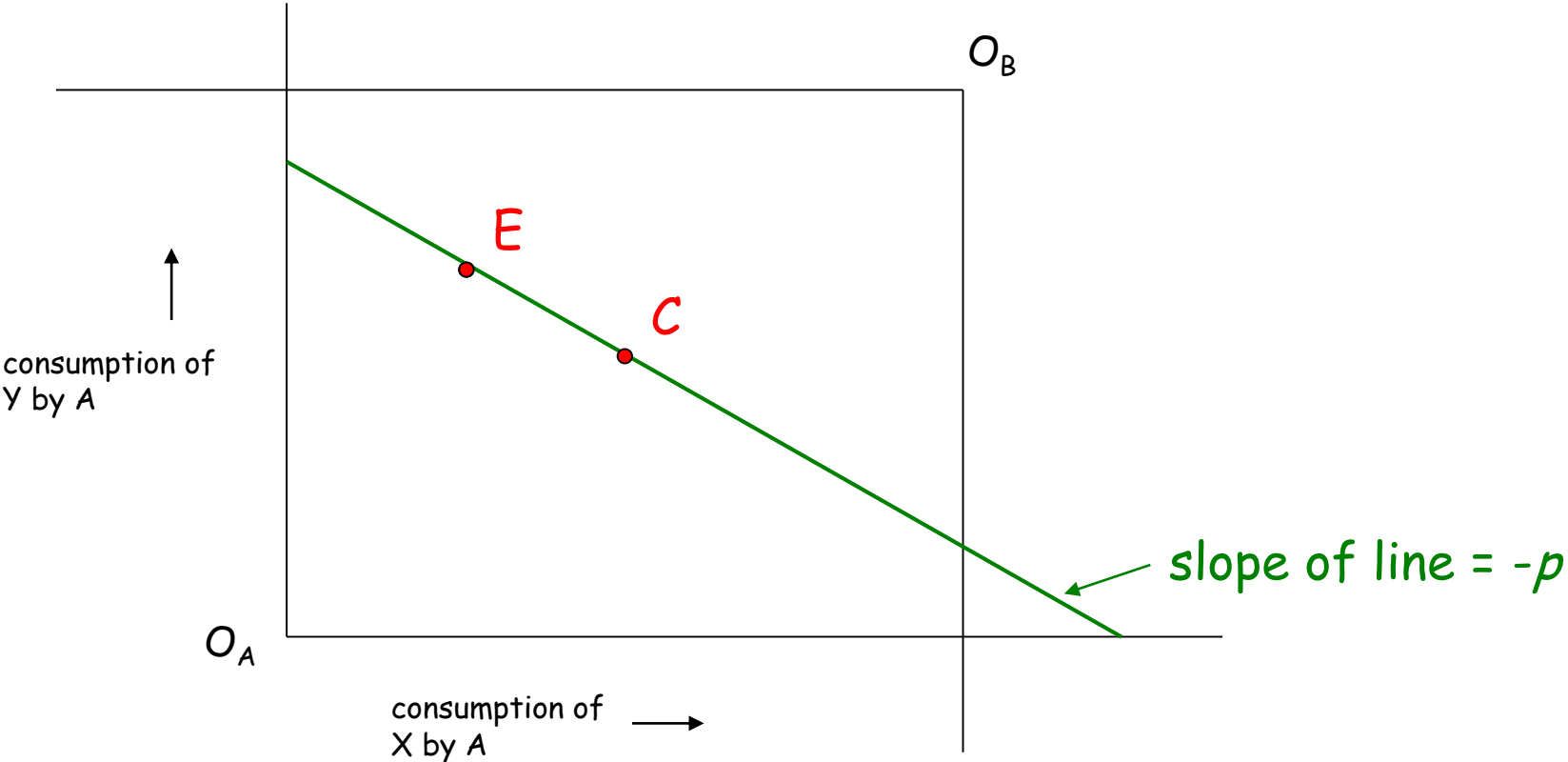
Now, suppose there is a single *market price* (p) of X in terms of Y, at which each individual can trade freely [can be positive or negative, because X may be good or bad];

and suppose that the final outcome of individuals' decisions is market-clearing (i.e. sales = purchases).

Notice that I am *not* assuming coherent preferences, only that each individual does *something* (or nothing) and that somehow A's and B's net sales sum to zero.

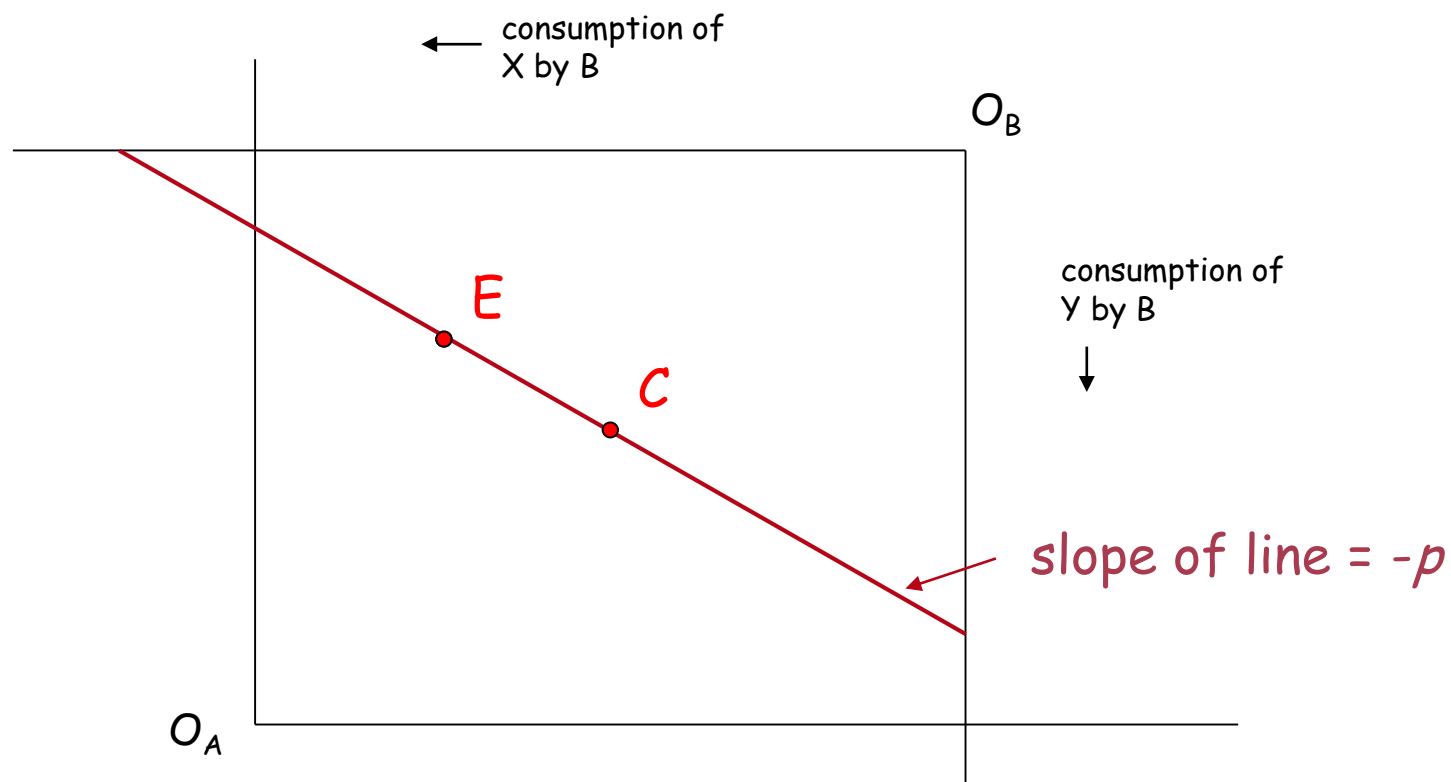
A's opportunity set is the green line.

His net trades take him from E (endowment) to C (consumption).



B's opportunity set is the red line.

Her net trades take her from E (endowment) to C (consumption); the market clears.



Opportunity criterion is satisfied iff: (i) individuals' actual choices from their opportunity sets are jointly feasible, and (ii) for every feasible transaction T among any given set S of individuals: *either* each member of S is free to choose his component of T *or* there is some member of S whose component of T is unambiguously worse than some other option he is free to choose.

Is this criterion satisfied at C ?

Part (i) is satisfied because C is jointly feasible. What about part (ii)?

A transaction that is feasible for $\{A, B\}$ must be a point in the Edgeworth box.

If the point is on the red/green price line, both individuals are free to choose it.

If it is below the line, it is dominated for A (there are options on the line which give the same amount of X and more Y).

If it is above the line, it is dominated for B (there are options on the line which give the same amount of X and more Y).

So the criterion is satisfied.