Economics and Philosophy, **34** (2018) 343–368 © Cambridge University Press
doi:10.1017/S0266267118000081 First published online 29 June 2018
cambridge.org/eap

ASSET INEQUALITY, ECONOMIC VULNERABILITY AND RELATIONAL EXPLOITATION

GILBERT L. SKILLMAN*

Abstract: In response to Roemer's reformulation of the Marxian concept of exploitation in terms of comparative wealth distributions (1982, 1996), Vrousalis (2013) treats economic exploitation as an explicitly relational phenomenon in which one party takes advantage of the other's economic vulnerability in order to extract a net benefit. This paper offers a critical assessment of Vrousalis's account, prompting a revised formulation that is analysed in the context of a matching and bargaining model. This analysis yields precise representations of Vrousalis's conditions of economic vulnerability and economic exploitation and facilitates comparison to the alternative conceptions of Marx and Roemer.

Keywords: inequality, exploitation, bargaining

According to the canonical Marxian account (Marx 1990), the existence of profit in capitalist economies is premised on the exploitation of labour by owners of alienable means of production, such that workers who are 'free in the double sense' (free to offer their labour power in the market, and 'free' of owning means of production) are compelled to expend more labour than is required to produce their wage bundles. This account of capitalist profit and exploitation thus features three key elements, involving characterizations of the class distribution of alienable productive assets (such that workers own no physical means of production), the economic relationship between capitalists and workers based on this distribution (involving market transactions for wage labour), and the form in which exploitation is manifested (capitalist extraction of surplus labour).

* Wesleyan University, Middletown, CT 06459, USA. URL: http://www.wesleyan.edu. Email: gskillman@wesleyan.edu

This account has been directly challenged by John Roemer in his seminal *A General Theory of Exploitation and Class* (1982) and related work (Roemer 1986, 1996: Part I). Roemer employs the analytical tools of general equilibrium theory to assess the coherence and generality of the Marxian account under alternative specifications respecting production technology, heterogeneity of labour inputs, and individual preferences for leisure and goods consumption. This analysis prompts Roemer to reject the traditional Marxian conception of exploitation and to advance a new formulation in which agents' exploitation status, measured in flows of welfare rather than labour, is determined by comparing existing economic outcomes with those attainable if wealth were equally distributed.[1] On the basis of this characterization, Roemer concludes that economic exploitation is best understood as a symptom of wealth inequality, so that normative concerns about exploitation ultimately translate as concerns about the justice of given distributions of productive assets.

In a critical response to this assessment, Nicholas Vrousalis (2013) proposes to treat exploitation as an explicitly relational phenomenon in which one party makes use of another party's vulnerability in order to extract a net benefit, and argues that this approach offers a robust foundation for the specifically Marxian conception of *economic* exploitation. I believe that this approach holds promise, but suggest that key aspects of Vrousalis's account of economic exploitation rest on unexamined claims regarding the systemic basis of economic power and the form in which such power is manifested. My purpose here is therefore to assess the economic and strategic logic of Vrousalis's account, with the intent of clarifying and refining his relational conception of exploitation.

This analysis proceeds in four steps. In the first section of the paper, I summarize Vrousalis's account of economic exploitation and his related notions of economic vulnerability and economic power. The critical point developed here is that a stipulation of strategically *credible* behaviour is necessary to establish an analytically coherent link between wealth inequality and economic vulnerability in the context of mutually voluntary exchange. In Section 2, I use a strategic bargaining model to show how a particular notion of strategically credible behaviour, reflected in the concept of *subgame-perfect* equilibrium, can be used to generate concrete hypotheses concerning the determinants of economic vulnerability. This analysis yields a precise connection between wealth differentials and the manifestation of economic vulnerability and exploitation in market transactions between capital and labour suppliers.

[1] In a subsequent reformulation, Roemer (1996: Essay 5) adds the stipulation that the exploiting coalition benefits from the labour performed by the exploited coalition. I discuss implications of this stipulation below.

Section 3 then extends the analysis to allow for the possibility of voluntary exit from given bargaining relationships, and discusses how the presence of exit options modifies the connection between wealth inequality and economic exploitation in the simple bargaining game. Section 4 further extends this analysis by considering the existence of exploitation within the specific scenario of Walrasian equilibrium, derived as the limiting case as exchange frictions approach zero of labour or capital market exchange characterized by matching and bargaining with voluntary exit (adapted from Rubinstein and Wolinsky 1985). The analysis here highlights the necessity of *capital scarcity* for the presence of exploitation under competitive conditions. After a closing discussion that compares and contrasts Vrousalis's account of economic exploitation with those of Marx and Roemer, proofs of propositions stated in the main text are presented in the Appendix.

A key difference in the formal analytical approach of this paper from that employed by Roemer involves the use of a *strategic*, rather than an *axiomatic*, game-theoretic framework. The strategic approach has both comparative benefits and drawbacks. On the plus side, the approach allows one to derive endogenously the connection between given structures of economic interaction and the existence and degree of exploitation, rather than simply stipulating these connections *a priori*. Relatedly, this approach makes it possible to distinguish the distribution of given gains from economic interaction relative to the anterior distribution of wealth, thus enabling a focus on the *relational* nature of exploitation.

On the downside, the connection between economic structures and exploitation depends on the specific structure of the bargaining model under study, so that the analytical results may lack generality. However, the outcomes yielded by the basic bargaining model studied here have some claim to robustness, as they correspond to the Nash bargaining solution, first derived on axiomatic grounds. This outcome is also generated as exchange frictions disappear in the Rubinstein and Wolinsky framework (1985) on the basis of a very different bargaining model.

1. VROUSALIS'S RELATIONAL NOTION OF ECONOMIC EXPLOITATION: CONSTITUENT ELEMENTS AND CLAIMS

Vrousalis advances a general definition of exploitation as a relational phenomenon which he then tailors to the specific case of *economic* exploitation. I'll start with his general definition, as it raises issues that will prove to be of particular relevance for the more exacting notion of economic exploitation that he ultimately treats.

Vrousalis posits that 'A exploits B if and only if A and B are embedded in a systematic relationship in which … A instrumentalizes …

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

B's vulnerability to extract a net benefit from B', where 'A and B are free variables referring ... to agents or groups or coalitions of optimizing agents' (Vrousalis 2013: 132). Thus, his relational conception of exploitation rests on the idea of one agent or group of agents being able to take advantage of the 'vulnerability' of another. I'll begin the discussion of Vrousalis's account by examining his characterization of this condition.

1.1. Relational vulnerability, credible threats and 'power over'

Vrousalis understands some individual B to be *vulnerable* 'just when he suffers a substantial risk of a significant loss in the relevant metric (welfare, resources, capabilities, and so on)' (133; parenthesis original). Vulnerability thus has two dimensions, respectively involving a loss and the prospect of incurring it, such that the potential loss is *significant*, and the risk of incurring it is *substantial*. Let's call B's potential loss a *dire prospect* if it meets both of these conditions.

Vrousalis depicts this condition of vulnerability as *relational* (as opposed to *absolute*) if B's dire prospect depends in an 'essential' way on another agent or set of agents A, such that A 'has some sort of power over B' (133). He posits in turn that A has power over B if and only if A is able to get B to do something for reason S (affected by a choice available to A) and B would not otherwise have chosen that action for reason S (136). In Vrousalis's formulation, then, B is relationally vulnerable to A if and only if B faces a dire prospect that somehow hinges on an action made by A, A has discretion with respect to taking that action, and B's consideration of A's discretionary action induces B to do something willed by A.

Vrousalis illustrates the condition of relational vulnerability and its connection to A's power over B with the scenario of *Deep Pit*, in which B is in a disadvantaged position relative to A, but one that A can easily ameliorate. Specifically, A and B are stuck in a deep pit subject to dangerous mudslides. B is stuck further down than A and is thus more vulnerable to mudslides, but A has a rope which can be extended to help B move up in the pit at no cost to A (133–4).

Vrousalis argues that 'B is vulnerable in the relational sense' in this scenario, and on the basis of this case asserts that a sufficient condition for B being relationally vulnerable to A obtains if (i) B 'lacks some desideratum x that is a requirement for, or a constitutive feature of, B's flourishing' (ii) which he can only obtain from A and (iii) which A 'has it within his discretion' to withhold from B (134). Label this condition $RV_{B,A}$. Do components (i)–(iii) of $RV_{B,A}$ collectively ensure that B suffers a substantial risk of a significant loss by virtue of B's relation to A?

My first major comment on Vrousalis's relational conception of exploitation is that scenarios similar to *Deep Pit* do not evidently establish a sufficient condition for B's being vulnerable to A. Condition $RV_{B,A}$

operationalizes the notion of 'significant loss' by contrasting B's condition with and without desideratum x, but it doesn't provide any basis for believing that B faces a risk, let alone a *substantial* one, that A would exercise his discretionary capacity to withhold what B needs. To infer this, we would also need to know something about A's *willingness* to withhold assistance from B.[2] Without this information, we consequently cannot know whether B is relationally vulnerable with respect to A, by Vrousalis's own specification. Furthermore, absent information about B's *beliefs* regarding A's willingness to withhold, and about B's choices in response to those beliefs, it cannot be inferred that A has power over B in such situations.

To illustrate these points, suppose we add to Vrousalis's *Deep Pit* scenario the condition that A would suffer a significant loss in a relevant metric were he *not* to provide assistance to B. Suppose, for example, that A has very strong feelings of empathy toward B, or has a well-developed sense of guilt that would be painfully engaged if he didn't extend the rope to B. Then it is at best not evident that B is relationally vulnerable to A in Vrousalis's sense. If it were also supposed that B knows about A's feelings, and knows, or at least believes, that they will compel A to provide the help that B needs, then it is even less evident that A enjoys power over B.

To anticipate the subsequent discussion of *economic* vulnerability and relational power, suppose alternatively that the *Deep Pit* scenario were augmented so that B also has something vitally needed by A that would be destroyed if B were buried in a mudslide, and which B has the discretion to withhold from A. A would thus face (at least) the *same* risk of significant harm if he were not to assist B, and would thus be relationally vulnerable. Suppose further that these facts are common knowledge to A and B. Any reason adduced to suggest that A has power over B would then evidently work just as well to establish that B had power over A.

These considerations indicate that B's relational vulnerability to A requires something more than just A's *discretion* to withhold a desideratum from B. In addition, A must be willing to act on this threat to withhold. Furthermore, in order for B's vulnerability to translate into A's power over B, it must also be the case that B *believes* that A is willing to exercise this discretion opportunistically, which is to say that B considers A's threat to withhold to be *credible*, in a sense to be discussed in the next section.

The conditions for securing power on the basis of credible threats are arguably even more strenuous when A is defined as a *group* and A's power over B requires that the individual members of A *collude* in exercising their discretion to harm B either through direct action or by withholding

[2] Indeed, Vrousalis acknowledges this in his account of *Deep Pit* when he says that 'B's vulnerability becomes a function of A's willingness to throw the rope' (134).

something that B needs. In this context, Vrousalis's assumption that coalitions are comprised of 'optimizing agents' does not suffice to ensure the agents' willingness to collude. For example, if the relationship among members of A has the payoff structure of a one-shot prisoners' dilemma, then although collusion is *collectively* optimal for members of A, it does not constitute a Nash equilibrium for the game.

Vrousalis treats the question of how power arises in relations among groups or coalitions in his *Three Thugs* scenario, in which he asserts a necessary and sufficient condition for any member of A (a trio of thugs) to have power over B (their intended mugging victim) 'is that the others be present and, if instructed, take B down' (143). However, Vrousalis does not identify conditions under which (1) the accomplices would be willing to act as 'instructed', and (2) B has any compelling reason to fear that the other thugs would act in this way, even if B believed that each thug were a rational optimizer. Consequently, this example does not suffice to establish B's relational vulnerability to group A, because it does not establish that the thugs can credibly threaten to collude in taking B down even if their collusion were collectively optimal.

1.2. Economic vulnerability and economic exploitation

Tailoring his general conception of exploitation to the specific context of economic relations, Vrousalis posits that 'A economically exploits B if and only if A and B are embedded in a systematic relationship in which … A instrumentalizes B's economic vulnerability to appropriate (the fruits of) B's labor' (138; parentheses original), where 'B is economically vulnerable to A if and only if B is vulnerable in virtue of B's position relative to A in the relations of production' understood as '… systematic relations of *effective ownership* … over human labor power and means of production in society' (136; emphasis original). Comparing this statement to Vrousalis's general definition of relational exploitation, it is clear that he understands '(the fruits of) B's labor' to be the *particular form* of the 'net benefit' extracted from B in such relations.

Before considering the implications of the 'net benefit' clause and its asserted manifestation in terms of B's labour, I want to discuss the specifically economic sense of relational vulnerability that Vrousalis invokes here. My second major comment on Vrousalis's relational approach to exploitation is that this depiction of *economic* vulnerability fails to distinguish relations of production that are mutually voluntary at the level of individual transactions from those that involve slavery or other forms of non-economic compulsion. For example, production relations between slaves and their owners certainly involve 'systematic relations of *effective ownership* … over human labor power and means of production', but they differ crucially from capitalist relations of production in that owners of means of production *also* effectively own the labour power of their slaves and can therefore exploit them by means presumably unavailable to individual capitalists.

As a consequence, Vrousalis's definition of economic exploitation, unlike Marx's or Roemer's, does not provide a basis for distinguishing pre-capitalist and capitalist forms of exploitation in relations of production. This lacuna is reflected in Vrousalis's contention that A's position vis-à-vis B in the *Deep Pit* scenario is 'exactly analogous to [A's having a] superior position in terms of wealth', so that 'economic vulnerability is a form of relational vulnerability such that, if B does not own any means of production (or, more broadly, wealth) and A does, or B owns substantially less than A then B is economically vulnerable to A and A has economic power over B' (137). The key difference is that in capitalist production relations, unlike in the *Deep Pit* scenario, the participation of all parties is voluntary.

In light of this observation, I suggest that Vrousalis's depiction of *relations of production* be modified to include the proviso that the relations in question are mediated by *mutually voluntary* market transactions between owners of labour power and owners of means of production. This modification is, at minimum, consistent with Vrousalis's intention to provide a counterpart to Roemer's notion of *capitalist* exploitation, which, like Marx's, presumes that economic exploitation arises on the basis of mutually voluntary individual transactions. One implication of this condition is that, since there are typically mutual prospective gains from economic interactions, it is possible that both parties are economically vulnerable. Consequently, something beyond mere wealth inequality must be posited in order to ensure that economic vulnerability is strictly unilateral in given economic transactions.

Even with this added stipulation, however, we encounter a potential divergence in Vrousalis's and Roemer's understanding of the systemic basis of economic exploitation. Vrousalis suggests that wealth inequality gives rise to exploitation because '... the wealth owned by capitalists systematically gives them a decisive *bargaining advantage* over workers, which means capitalists always have and can take advantage of economic power over workers (and never vice versa)' (137; emphasis added). In Roemer's framework, however, exploitation arises even though all actors are price takers and thus have no bargaining power. This is also broadly consistent with Marx's characterization of capitalist competition (Marx 1990: 1014; 1991: 275). In the following sections, I address this apparent divergence in the two accounts by placing economic transactions in the context of a matching and bargaining framework in which competitive market conditions emerge in the limit as exchange frictions approach zero. This approach provides a common terrain for assessing the alternative ways in which wealth inequality might convey economic power.

Finally, return to Vrousalis's stipulation that the 'net benefit' received by the exploiter in relations of production is manifested as (the fruits of) the labour of the exploited party, for which he offers no explicit justification. This would presumptively be the case in the canonical Marxian scenario of relations of production between wealthy capitalists and workers 'free in the double sense', since the latter would have nothing to offer but their ability to labour. However, Vrousalis's definition of economic vulnerability does not require that workers are entirely 'free' of owning their own means of production, so that the net benefit received by the agent with economic power might conceivably be based on something other than, or in addition to, the other's labour contribution. For example, a poor but not property-less worker might be induced to convey a net benefit to a wealthier capitalist by (also) foregoing current consumption possibilities to enable increased investment in productive capacity. Consequently, I suggest that the form of the net benefit received by the exploiting agent is best treated as an inference from given exchange conditions, rather than as an *a priori* requirement.

1.3. The 'net benefit' from exercising economic power: what's the counterfactual?

In Vrousalis's account, B's economic vulnerability to A leads to economic *exploitation* just when A instrumentalizes that vulnerability to gain a *net benefit*. But 'net' relative to what alternative? My final major comment on Vrousalis's relational account concerns the counterfactual scenario by which one might determine possible gains by the party enjoying economic power.

Vrousalis does not precisely specify the relevant counterfactual for judging B's exploitation status. However, his conception clearly differs from Roemer's in that it does not focus on wealth inequality *per se*, but rather on the implications of such inequality for the poorer agent's economic vulnerability. Specifically, while Roemer's assessment of capitalist exploitation is based on a hypothetical alternative in which alienable assets are equally distributed, Vrousalis's conception of economic exploitation inquires only about B's prospects in the event that A chooses not to transact with B. According to $RV_{B,A}$, B 'lacks some desideratum x (possessed by A) that is a requirement for, or a constitutive feature of, B's flourishing'. This says nothing about A's outside prospects other than that A possesses x. More generally, B is said to be vulnerable 'just when' he faces a dire prospect outside of the relationship with A.

This suggests that the 'net benefit' received by an exploiter should be assessed relative to a counterfactual situation in which B is able to flourish, or at least to avoid a dire prospect, outside of the relationship with A. Note that insofar as production relations typically yield a net surplus that can be shared among participants, this hypothetical alternative would not rule out the possibility of mutually beneficial and non-exploitative transactions between A and B. This is an important point of divergence of Vrousalis's account of economic exploitation from that of Marx, which maintains that *any* surplus received by capitalists from employing workers is based on the exploitation of their labour power.

1.4. Assessment

The foregoing discussion raises no fundamental objections to Vrousalis's general conception of exploitation, as such. It does, however, suggest that the relevance and coherence of all the major elements of Vrousalis's conception of economic exploitation hinges on the meaning, basis and implications of relational vulnerability in economic transactions. I've developed three points in this connection:

(1) The condition that B is relationally vulnerable to A requires something more than A's *discretion* to withhold something valued by B. For B to face a dire prospect, it must also be established that B regards a threat by A to withhold the good as *credible*.
(2) Vrousalis's depiction of *economic* vulnerability should be modified to reflect the fact that economic transactions are mutually voluntary, and thus may yield gains to all transactors. This raises the possibility that all parties to a given transaction are economically vulnerable, so that additional conditions are needed to ensure that economic vulnerability, where it arises, is unilaterally incurred. Under such conditions, wealthier agents may enjoy asymmetric economic power, but it need not follow that the 'net benefit' they enjoy takes the form of the exploited's labour or its fruits unless workers own no alienable wealth.
(3) Assessing the 'net benefit' received by an exploiter requires a determination of economic outcomes obtainable in the absence of economic vulnerability. Unlike in Roemer's conception of capitalist exploitation, the relevant counterfactual state in Vrousalis's account need not involve the equalization of alienable wealth. Unlike in Marx's account, it need not preclude A from realizing economic gains from a non-exploitative interaction with B.

2. STRATEGIC CREDIBILITY AND ECONOMIC POWER IN INDIVIDUAL CAPITAL-LABOUR TRANSACTIONS

In this section, I use a game-theoretic bargaining model to illustrate how considerations of strategic credibility might inform the link between wealth inequality and economic power in a representative exchange relationship involving production. After introducing the model and

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

tailoring the conception of economic vulnerability previously discussed to this framework, I characterize bargaining equilibrium and discuss its implications for the existence of economic vulnerability and exploitation.

2.1. Agents, preferences and payoff possibilities

In keeping with Vrousalis's focus on relations of production involving effective ownership over means of production and labour power, let there be two types of agents, denoted by K and L and indexed by subscript i.[3] Let subscript $-i$ denote 'the agent other than i' and let subscript j be used to denote a specific agent. Suppose that agents are infinitely lived, with time advancing in discrete steps indexed by t. Let τ be used to indicate a given time period. In this section, I'll consider a bargaining relationship between a representative pair of K- and L-type agents.

Suppose that the payoffs of type-i agents at any period τ are given by the present discounted value of income flows from that time forward, expressed as $\pi_i = \sum_{t=\tau}^{\infty} \delta^{t-\tau} y_{it}$, where y_{it} denotes the net income received by an agent of type i in period t and $\delta \in (0, 1)$ represents the time discount factor common to all agents. I'll discuss the implications of differential rates of time preference later in the argument.

Each agent of type i has an endowment of productive assets generating an income of $a_i > 0$, $i = K, L$, with a corresponding present discounted value of lifetime autarkic income, starting from any period τ, given by $A_i = \sum_{t=\tau}^{\infty} \delta^{t-\tau} a_i = a_i/(1-\delta)$. To capture the idea that K-type agents have greater wealth, assume that $A_K > A_L$.

Now suppose that the production relationship ensuing from a successful pairing of K- and L-type agents yields a value per period, net of input costs, equal to $v > 0$. The *surplus* generated by this production relationship relative to what the agents could otherwise secure for themselves is thus given by $s = v - (a_K + a_L)$. I'll assume that any pairing of K- and L- type agents generates a strictly positive surplus, so that $v > (a_K + a_L)$. Suppose that a given production relationship, once commenced, continues forever, and let the corresponding discounted present values of an infinite stream of v and s per period be denoted respectively by V and S.

2.2. The bargaining process

Suppose that a given pair of K and L agents determine the distribution of prospective production income V between them by an alternating-offer bargaining process similar to that studied by Rubinstein (1982). Let the agent making the initial offer be determined by the toss of

[3] Where there is no risk of confusion, I shall also use K and L to refer to representative agents of each type.

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

a fair coin,[4] after which agents alternate in making offers as long as bargaining continues. In each period, the agent receiving an offer can respond in either of two ways: accept the offer, in which case bargaining immediately concludes forever and the agents share the payoffs according to the accepted proposal, or reject the offer, in which case both agents immediately receive their respective autarkic payoffs[5] and the rejecting agent makes a counter-offer in the following period. Bargaining continues until some agent's offer is accepted; if this never occurs, the agents' payoffs are their respective discounted autarkic income streams(A_K, A_L). Note that since agents have the option of rejecting the other's offers forever, each agent $i = K, L$ can secure a payoff of *at least* A_i and *at most* $V - A_{-i}$.

Following Rubinstein, agents' bargaining costs per period are represented by their discount factor δ, reflecting the agents' common degree of time preference or 'impatience'. By choosing to reject a standing offer, an agent thus incurs the utility cost of delaying by at least a period the distribution of prospective gains from the production relationship.

Note that the bargaining game has a stationary structure after the initial coin toss to determine who makes the first offer. Every two periods, the same subgame arises, with the same order of moves, number of remaining periods, autarkic payoffs, and surplus to be distributed. As will become clear, this stationarity plays a central role in the derivation of equilibrium and the specification of economic vulnerability in the context of the bargaining relationship.

A *strategy* for a given agent i specifies what action that agent would take at each point in the bargaining game that he or she is called upon to make a move (i.e. to make an offer or respond to one with acceptance or rejection), contingent where relevant on the moves chosen by the other player. Let Z_i represent the set of possible strategies for player i, and correspondingly let $Z = Z_K \times Z_L$ denote the set of all possible strategy *combinations* (z_K, z_L).

The bargaining game is therefore completely described by the value to be shared (V) plus the players' discount factor δ and possible strategy combinations Z. Denote this game by Γ. Let a given pair of equilibrium

[4] I introduce this assumption to distinguish other sources of bargaining power from the well-known 'first-mover advantage' enjoyed by initial proposers in a sequential bargaining process (demonstrated in the proof to Proposition 1 included in the appendix). This assumption is not innocuous if it were the case that superior wealth somehow conveys the power to make initial offers.

[5] This is a significant departure from the standard Rubinstein model, which assumes, arbitrarily, that both players receive a payoff of zero in any period that agreement is not reached. This modification is, indeed, central to the equilibrium connection subsequently derived between inequality and exploitation. I thank an anonymous referee for pressing me to explore this modification of the Rubinstein bargaining model.

payoffs for the bargaining subgame in which player $j = K \text{ or } L$ makes the first offer be expressed as(π_K^{j*}, π_L^{j*}), and let the corresponding *expected* equilibrium payoffs for the bargaining game be given by (Π_K^*, Π_L^*), where $\Pi_i^* = (\pi_i^{K*} + \pi_i^{L*})/2$, $i = K, L$, the probability-weighted average of the subgame-specific payoffs.

2.3. Strategic credibility, economic vulnerability and exploitation

What are the possible equilibrium outcomes of this game? To answer this question, we need to say something about the admissible strategic choices of the agents. Vrousalis allows that agents are 'optimizing' (2013: 132), but the assumption of optimizing behaviour will generally not be sufficient of itself to determine how the bargaining game would be played in theory.[6]

To see this, suppose that we were to adopt the basic solution concept for non-cooperative or strategic games, that of *Nash equilibrium*. If z_i denotes a bargaining strategy for player i, then a given pair of strategies (z_K^N, z_L^N) constitutes a Nash equilibrium for the bargaining game if, for each agent i, z_i^N maximizes his or her payoff, *given* the equilibrium strategy of the other player. The concept of Nash equilibrium thus clearly incorporates the assumption of individually optimizing behaviour, along with the requirement that the players' optimizing behaviours are mutually consistent in a non-cooperative sense.

However, the assumption that agents' strategic choices correspond to a Nash equilibrium doesn't tell us much about how the bargaining game is played or how economic power is determined. That is because *any* distribution of the available surplus such that $\Pi_i \geq A_i \; \forall i$ (known as the 'individual rationality' condition) can be supported by Nash equilibrium strategies. Suppose, for example, that $(\hat{\Pi}_K, \hat{\Pi}_L)$ is a pair of payoffs which satisfy the individual rationality condition and sum to the total value V. This distribution can be supported by a pair of strategies such that each player i insists on $\hat{\Pi}_i$ when it is her turn to make an offer, and accepts any standing offer if and only if it yields her at least $\hat{\Pi}_i$. This point is of critical significance to Vrousalis's account of economic power and exploitation, as it shows that some assumption beyond that of optimizing behaviour is necessary in order to show how economic power is derived from differences in individual economic circumstances.

6 The sense of 'optimizing behaviour' that I apply here is one of *self-regarding* optimization, such that each player looks only to maximize his or her own expected payoff, given the behaviour of the other player(s). This is the reasoning by which subgame-perfect equilibria for strategic bargaining games have typically been derived. However, one could entertain more 'enlightened' notions of optimization that, for example, allow for considerations of fairness or other norms. Doing so would make the issue of whether would-be exploiters choose to 'instrumentalize' their economic power less trivial than is indicated below.

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

Rubinstein (1982) proposed to resolve this fundamental indeterminacy by assuming that bargaining occurs over time and requiring that bargaining strategies satisfy the condition of *subgame-perfect equilibrium*, defined as a pair of strategies whose elements constitute a Nash equilibrium for *every subgame* of the overall game Γ. The force of this requirement is that players' strategic choices must be *credible* in the sense that commitments to *future* moves are in fact optimal to carry out at the time that the player must act on such commitments. Suppose, for example, that a particular bargaining strategy compels a player at a given stage in the game to make a counter-offer rather than accepting a standing proposal, even though the expected payoff from doing so is lower than that provided by the standing offer. The requirement that admissible solutions satisfy subgame-perfect equilibrium would rule out making a counter-offer at this point in the game as a component of the agent's equilibrium strategy.

I will adopt the condition of subgame perfection in the following analysis in order to explore the implications of strategic credibility for the determination of economic vulnerability and economic power, without insisting that this is the only relevant or coherent way to incorporate this condition into the analysis of economic exploitation.[7] It should be clear from the foregoing, however, that *some* notion of strategic credibility beyond the dictates of Nash equilibrium is needed to establish a tight connection between wealth inequality and exploitation.

In light of the conceptual discussion in the previous section, how might Vrousalis's relational notions of economic vulnerability and exploitation be represented in this context? Define $A^f > 0$ as the present discounted value of lifetime income minimally necessary for any agent to flourish. First, assume that $A_K > A^f > A_L$, so that only agents of type L face the prospect of not flourishing in autarky, and thus that economic vulnerability, if it arises, is unilateral. Second, to ensure that any vulnerability of type-L agents flows from the economic relationship with K-type agents rather than simply reflecting the latter's superior wealth, assume that $V - A_K > A^f$.

Next, recall that Vrousalis's vulnerability condition $RV_{B,A}$, as amended in the previous section, posits that K can *credibly* withhold a benefit that L needs in order to flourish. In the context of the bargaining relationship described above, this case holds in equilibrium if either K can credibly reject any offer that leaves L a residual of at least A^f or L cannot credibly refuse offers below A^f. Given the stationary structure of the bargaining game, these conditions will hold in equilibrium in the first round of the bargaining game if they hold in any round.

[7] A more general notion of credible behaviour would be needed, for example, if the possibility of irrational play were incorporated (Binmore 2007: Appendix B).

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

In light of these considerations, posit that an agent of type L is *economically vulnerable* in a bargaining relationship with a type-K agent if, in any subgame-perfect equilibrium, either the latter can credibly reject any offer less than or equal to $V - A^f$ *or* the type-L agent *cannot* credibly reject some offers less than A^f. Since each agent has a 50% chance of making the first offer, agent L faces at least that probability of receiving a payoff below A^f if either condition holds, which prospect is correspondingly interpreted as a dire prospect for L. Finally, posit that K *exploits* L if K's equilibrium payoff is higher than the *maximum* equilibrium payoff that K could receive if L were not economically vulnerable in the sense just defined.

2.4. Equilibrium bargaining outcomes and exploitation

The requirement that admissible strategies be subgame-perfect has considerable power in restricting the range of equilibria attainable in the simple bargaining game described above. Indeed, this game has a unique equilibrium with properties summarized in the following proposition. (Proofs of all propositions are presented in the Appendix.)

Proposition 1. *There is a unique subgame-perfect equilibrium in which bargaining concludes immediately and agents split the production value V. In this equilibrium, each agent's expected payoff is strictly increasing in his or her autarkic income and strictly decreasing in the autarkic income of the opposing agent.*

This proposition corroborates Vrousalis's contention that economic inequality conveys a bargaining advantage to those with superior wealth (2013: 137), given the assumption that autarkic income flows mirror wealth differences. More specifically, the expected payoffs described in the proposition correspond to the well-known *Nash bargaining solution* of cooperative game theory for a game in which V is the total value to be shared and (A_K, A_L) are the 'disagreement' payoffs.[8]

Now consider the implications of subgame-perfection for agent L's economic vulnerability and exploitation status.

Proposition 2. *Agent L is economically vulnerable to agent K if and only if* $[\delta(V - A_K) + A_L]/(1 + \delta) < A^f$. *In that case, L is economically exploited by K.*

Just as with the determination of equilibrium payoffs, the economic vulnerability condition is driven by the restriction to subgame-perfect

[8] See Binmore (1986) for a similar result. The Nash bargaining solution is also generated as a limit case of the bargaining model studied by Rubinstein and Wolinsky (1985) via an essentially different process, in which existing relationships are terminated with exogenously given probabilities after each round in which agreement is not reached.

strategies. As shown in the proof of Proposition 1, the unique equilibrium is such that the type-L agent can only expect to receive a present-value payoff of $[\delta(V - A_K) + A_L]/(1 + \delta)$ upon rejecting a standing offer and making a counteroffer in the following period, and thus can only credibly reject offers less than that amount. If this value falls below the present discounted value of the income stream L needs in order to flourish, then L is economically vulnerable by the definition given above.

One can also see from the proof of Proposition 2 that if it were the case that $[(V - A_K) + \delta A_L]/(1 + \delta) < A^f$ (a stronger condition, since $(V - A_K) > A_L$ and $\delta < 1$ by assumption), then agent K can credibly reject offers which yield a residual of at least A^f, so that L cannot expect to flourish no matter who is selected by chance to make the first offer. In any case, the result that L is exploited under the condition stated in Proposition 2 follows from the conditions that L is vulnerable and K's expected payoff is strictly decreasing in A_L.

An additional basis for economic exploitation would arise in the case that there is a systematic connection between agents' autarkic wealth levels and their respective bargaining costs. The expected bargaining payoffs reported in Proposition 1 are obtained under the assumption that agents' discount factors are parametric (and equal) and thus have no systematic connection to their endowments. Suppose instead that agents' time preferences are identical but wealth-elastic, implying that individual discount factors are increasing in individual wealth, and assume correspondingly that $\delta_K > \delta_L$. There is some evidence for this hypothesis of *diminishing marginal impatience* (DMI) (see Lawrance (1991) and Samwick (1998) for econometric evidence and Harrison *et al*. (2002) for experimental evidence in support of this hypothesis).

In this case, inequality in outside payoffs also affects agents' bargaining costs, such that the per-period cost of delay is always higher for agents with relatively low wealth levels. This has two consequences for the exploitation status of L-type agents in the present model. (Propositions and proofs relating to these claims are omitted here, but available from the author upon request.) First, relative to the case in which both agent types have the same discount factor, DMI expands the conditions under which agents with lower wealth are economically vulnerable for given wealth distributions by raising their cost of delayed agreement. Second, DMI raises the equilibrium expected payoff to the wealthier agent, and thus increases the degree of economic exploitation in the case that L-type agents are vulnerable.

3. ECONOMIC EXPLOITATION WITH EXIT OPTIONS

Bargaining in the presence of market competition introduces the possibility that parties to an economic transaction might respond to an unsatisfactory offer by exiting the bargaining relationship instead of electing to continue the bargaining process for at least another period. To assess the implications of such *exit options* for the manifestation of economic vulnerability and exploitation, suppose that the bargaining process discussed in the previous section is augmented to allow the responding player in any period either to reject the offer and make a counteroffer in the following period, as before, or else exit the relationship entirely.

3.1. The bargaining game with exit options

Let the respective expected payoffs to agents K and L from exiting a given bargaining relationship be given by (W_K, W_L), and assume once again that all agents have the parametric discount factor $\delta \in (0,1)$. In the absence of specific exit barriers, it is reasonable to suppose that $W_i \geq A_i$, since agents could elect to leave the market entirely. However, the exit payoffs of given agents will in general also depend on their prospects for engaging replacement exchange partners and the payoffs that might be expected from these alternative transactions. Endogenous determination of exit payoffs will be discussed further below; for now, exit payoffs are taken as parametric and at least as great as agents' autarkic income flows. It is convenient to assume that $V \geq W_K + W_L$; later it will be shown that this inequality must hold in market equilibrium.

3.2. Equilibrium payoffs and exploitation in the bargaining game with exit options

The following proposition characterizes equilibrium payoffs for the bargaining game with exit.

Proposition 3. *The bargaining game with exit options has a unique subgame-perfect equilibrium in which initial offers are immediately accepted and the agents split the production value V. There are three equilibrium scenarios:*

- *(E1) If both agents' outside payoffs are no greater than what they could respectively secure by rejecting the opponent's offer in the bargaining game without exit, then each agent's equilibrium payoff is as described in Proposition 1.*
- *(E2) If both agents' outside payoffs are strictly greater than what they could respectively secure by rejecting the opponent's offer in the bargaining game without exit, then each agent's equilibrium payoff is strictly increasing in his or her own outside payoff and strictly decreasing in the opposing agent's outside payoff.*
- *(E3) If one agent's outside payoff is relatively high in the sense described in (E2) and the other agent's outside payoff is relatively low in the sense described in (E1), then the payoff of the first agent is strictly increasing in his or her*

outside payoff and strictly decreasing in the autarkic income of the latter player, while the expected payoff of the other agent is strictly increasing in his or her autarkic income and strictly decreasing in the exit payoff of the opposing agent.

Given the restriction to subgame-perfect bargaining strategies, an agent will only elect to exit in response to an unsatisfactory offer if doing so yields a payoff at least as high as that which can be secured by rejecting the offer but staying in the existing relationship. Consequently, exit payoffs can only effect equilibrium bargaining outcomes if they are sufficiently high relative to the payoffs that would otherwise obtain in the absence of exit options. This is the logic underlying the three equilibrium scenarios reported in Proposition 3. Experimental and econometric evidence for the predicted contingent impact of exit options is provided respectively by Binmore *et al.* (1989) and Scaramozzino (1991).

A corollary of this observation is that the presence of credible exit options puts a floor on each agent's equilibrium bargaining payoffs. As a direct consequence, the scope of L's economic vulnerability is non-decreasing in L's own exit payoff and non-increasing in K's exit payoff, as specified in the following proposition. To facilitate the statement of the following result, let us say that a strategy is *strictly credible* if it yields a strictly higher payoff to a player than the next best feasible strategy.

Proposition 4. *For given A_K and A^f, K's exit option expands the range of values of A_L for which L is economically vulnerable if and only if K's exit option is strictly credible, and L's exit option contracts the range of values of A_L for which L is economically vulnerable if and only if L's exit option is strictly credible. In any case, L is economically exploited by K if L is economically vulnerable.*

In order to determine how the presence of exit options affects L's exploitation status in given market settings, we must understand the market conditions affecting the expected payoffs to exit of all players. This issue is taken up in the next section.

4. ECONOMIC EXPLOITATION IN (COMPETITIVE) MARKET EQUILIBRIUM

Now suppose that the bargaining relationship studied previously is embedded in a market process in which the K- and L- type agents are randomly matched with probabilities determined by their respective numbers in the market. This step makes it possible to link agents' exit payoffs endogenously to underlying market conditions as well as to autarkic income flows based on agents' respective wealth holdings. This matching and bargaining framework is studied by Rubinstein and Wolinsky (1985) using a bargaining model in which given transactions are assumed to be terminated exogenously with some positive probability rather than voluntarily chosen, as assumed in the previous section. A corresponding matching and bargaining framework with voluntary exit is analysed in Skillman (2016). I summarize the relevant implications of that analysis here, focusing primarily on the case of competitive equilibrium that emerges as matching frictions approach zero. The key issue under study concerns the conditions under which market competition modifies the bargaining connection between wealth inequality and economic exploitation through its impact on the determination of exit payoffs.

Following Rubinstein and Wolinsky (1985), the market mechanism pairing agents of different types is represented as a probabilistic matching process based on net agent flows per period. Let $N_{it} > 0$ represent the number of type-i agents in the market at time t. Then the number of matches of L and K agents in any period t is given by a *matching function* $M_t = m(N_{Kt}, N_{Lt})$, where M_t denotes the number of matches of K and L agents in period t and the function m is assumed to be non-decreasing in its arguments and $M_t \leq \min\{N_{Kt}, N_{Lt}\}$. The corresponding probability that an agent of type i will be matched with an agent of complementary type in period t is indicated by $q_{it} = m(N_{Lt},\ N_{Kt})/N_{it}$.

I depart from Rubinstein and Wolinsky, however, in specifying that the matching function takes the specific form $M_t = \min\{N_{it}, N_{jt}\}(1-\varepsilon)$, where $\varepsilon > 0$ is a parameter representing frictions in the matching process. I refer to this matching function as *semi-Walrasian* because it captures the asymmetry that, save for the presence of matching frictions, agents on the 'short side' of the market are certain of being matched.[9]

In any period τ, then, an unmatched agent has the probability $q_{i\tau}$ of being matched and engaging in bargaining with an agent of complementary type. With corresponding probability $1 - q_{i\tau}$, the agent remains unmatched, receiving an immediate payoff of a_i and then re-entering the matching process in the following period. Letting $W_{i\tau}$ denote the expected present payoff of an agent who is unmatched at the beginning of period τ, it follows that an unmatched agent has an expected payoff of $a_i + \delta_i W_{i,\tau+1}$.

Consequently, the value of an agent's exit option in a given period τ is expressed by $W_{i\tau} = q_{i\tau}\Pi^*_{i\tau} + (1-q_{i\tau})[a_i + \delta W_{i,\tau+1}]$, where $\Pi^*_{i\tau}$ denotes an equilibrium bargaining payoff for agent i in period τ and $W_{i,\tau+1}$ denotes

[9] The issue at stake here goes beyond the specification of the matching function, however. A key feature of Rubinstein and Wolinsky's analysis relates to their demonstration that a non-Walrasian equilibrium emerges from their model in the limit as matching frictions approach zero. This assessment is criticized by Gale (1987), who argues that Rubinstein and Wolinsky fail to establish exchange conditions sufficient to ensure steady-state numbers of agents in the market. Building such conditions into Rubinstein and Wolinsky's framework, Gale argues, yields Walrasian equilibrium outcomes in the limit. Thus, one might interpret the semi-Walrasian specification of the matching function as an alternative way of conveying Gale's insight.

the expected payoff to an agent of the same type who is unmatched at the beginning of the following period. Recall from Proposition 3, however, that bargaining payoffs depend in turn on the values of exit options in two equilibrium scenarios. Thus, to close the system and yield determinate outcomes, a condition is needed linking current to future exit payoffs.

For the overall matching and bargaining game, I follow Rubinstein and Wolinsky in limiting attention to *semi-stationary* strategies, which requires that (1) all agents of the same type play the same strategies, (2) agents of each type play the same strategy in any match, and (3) there is a *steady state* in agent flows such that for each i, $N_{it} = N_i$.

Now consider how agents' outside payoffs are determined by the market matching process in a semi-stationary equilibrium. Denoting the expected payoff of a newly matched agent by Π_i^s and the expected steady-state payoff of a currently unmatched agent by W_i^s, $i = K, L$, it follows that the value of the latter in a semi-stationary equilibrium is given by $W_i^s = q_i \Pi_i^s + (1 - q_i)[a_i + \delta_i W_i^s]$, implying in turn that the steady-state value of an unmatched player's expected payoff is expressed as

$$(1) \qquad W_i^s = [q_i \cdot \Pi_i^s + (1 - q_i) \cdot a_i]/(1 - \delta_i(1 - q_i)),\ i = K, L.$$

Expression (1) indicates that the expected payoff to an unmatched player depends on the probability-weighted average of the bargaining payoffs from being successfully matched and from autarkic income in a single period, where the probability weights are determined by the matching function and steady-state player flows. It thus illustrates the specific but partial link between outside payoffs and underlying wealth inequality.

The steady-state equilibrium payoff to a matched player i, $i = K, L$, is then determined by the condition that

$$(2) \qquad \Pi_i^s = \Pi_i^* \big|\, W_j = W_j^s,\ j = K, L,$$

implying that exit and bargaining payoffs are simultaneously determined in equilibrium.

In general, all three of the bargaining scenarios described in Proposition 3 can be sustained in the semi-stationary market equilibrium so long as matching probabilities and agents' discount factors are both sufficiently below one, with the added condition that outside payoffs are determined endogenously by autarkic income flows, bargaining payoffs and matching probabilities. However, for the present purpose of discussing Vrousalis's relational conception of exploitation, I want to focus on the conditions under which economic vulnerability and thus exploitation arise in the limiting equilibrium as the matching friction parameter ε approaches zero, understood as a scenario of 'competitive' market equilibrium.

Proposition 5. *Suppose that the matching function is semi-Walrasian and the numbers of both agent types are at their respective steady-state values. Then in the limit as matching frictions approach zero, agents on the long side of the market receive just their autarkic income flows, while agents on the short side of the market receive the entire residual. If there are equal numbers of K- and L-type agents in the market, equilibrium payoffs are indeterminate, such that the equilibrium payoff to an agent of a given type ranges from that agent's autarkic income to the entire production value V net of the other agent's autarkic income.*

Assuming that the matching function is semi-Walrasian, the matching and bargaining process assumed here mimics a Walrasian competitive equilibrium in the limit as matching frictions approach zero. Thus, agents on the short side of the market receive the entire surplus in the limit equilibrium, leaving agents on the opposite side with just their autarkic payoffs. Implications for the incidence of exploitation are stated in the final proposition.

Proposition 6. *In the limit as matching frictions approach zero under the conditions of Proposition 5, market equilibrium is such that agents of type L are exploited if they are on the long side of the market, but are not exploited if they are on the short side of the market. If there are equal numbers of each agent type, there are both exploitative and non-exploitative equilibria.*

Proposition 6 shows that having superior wealth need not translate into exploitative power when the market supporting capital-labour exchanges is competitive, because a surfeit of capital suppliers drives capital's share of the surplus to zero. A similar issue regarding the necessity of 'capital scarcity' for positive equilibrium exploitation arises in Roemer's analysis (1982: 9–11), with the added caveat that such scarcity can also be undone by large capital holdings of relatively few capital suppliers. By extension, the process of capital *accumulation* may serve to reverse the condition of capital scarcity over time. Viewed in this light, Marx's discussion of the 'industrial reserve army' concerns the persistence of capital scarcity, and thus of capitalist exploitation, in the presence of capital accumulation (Marx 1990: 781–94).

5. ALTERNATIVE APPROACHES TO ECONOMIC EXPLOITATION

As interpreted and analysed here, Vrousalis's relational conception of economic exploitation overlaps significantly with Marx's and Roemer's respective notions of capitalist exploitation, but is equivalent to neither of them. The overlap among the three accounts is most readily seen in the scenario of competitive equilibrium, as this is (allowing for variations in formal conception) the primary case studied by Marx and Roemer. By all three accounts, capital owners exploit suppliers of labour power given the presence of equilibrium *capital scarcity*, understood as a situation in which there is excess supply of labour power (i.e. an 'industrial reserve army'). Correspondingly, all three accounts would agree in finding no exploitation of labour power suppliers in the case that capitalist profits are zero (although this assessment is only assured in Roemer's account by his added proviso that capitalists gain from their relations with labour).

A potential basis for divergent assessments of economic exploitation by the three accounts emerges in the case that mobility frictions preclude attainment of competitive market conditions and allow scope for bargaining in the determination of profits and wages. For Marx, any outcome yielding positive profits would give rise to exploitation so long as it compelled workers to expend more labour time than is socially necessary to produce their wage bundles. For Roemer, the outcome would be exploitative given that workers could be made better off by an egalitarian redistribution of alienable productive assets. In contrast, Vrousalis's relational conception of exploitation stipulates the potentially more stringent condition that workers' bargaining position is such that they could not command a share of the production surplus that allows them to flourish, and so are economically vulnerable. However, as discussed in Section 3 above, capitalists' superior wealth does not of itself suffice to establish this condition.

It may of course be the case that this stipulation is typically met in real-world capitalist economies, at least for less-skilled labour, but ascertaining this would require a determination of worker's bargaining power in given markets, and more fundamentally, of the income levels minimally necessary to protect labour suppliers from a condition of economic vulnerability. While this is no trivial task, it is already being addressed in academic and policy debates concerning the provision of basic income or living wage guarantees to less advantaged members of society. Vrousalis's relational conception of exploitation can thus be thought of as providing a specific normative basis for such proposals, even as ongoing research on these matters provides theoretical and empirical substance for his conception.

ACKNOWLEDGEMENTS

With the standard caveat, the author thanks Roberto Veneziani, Nicholas Vrousalis, Naoki Yoshihara, and workshop participants at Queen Mary University of London and Tohoku University, Sendai, for helpful comments on earlier drafts.

REFERENCES

Binmore, K. 2007. *Does Game Theory Work? The Bargaining Challenge*. Cambridge, MA: MIT Press.

Binmore, K., A. Rubinstein and A. Wolinsky. 1986. The Nash bargaining solution in economic modelling. *Rand Journal of Economics* 17: 176–188.

Binmore, K., A. Shaked and J. Sutton. 1989. An outside option experiment. *Quarterly Journal of Economics* 104: 753–770.
Gale, D. 1987. Limit theorems for markets with sequential bargaining. *Journal of Economic Theory* 43: 20–54.
Harrison, G., M. Lau and M. Williams. 2002. Estimating individual discount rates in Denmark: a field experiment. *American Economic Review* 92: 1606–1617.
Lawrance, E. 1991. Poverty and the rate of time preference: evidence from panel data. *Journal of Political Economy* 99: 54–77.
Marx, K. 1990 [1867]. *Capital: A Critique of Political Economy*. Volume I. Harmondsworth: Penguin Books.
Marx, K. 1991 [1894]. *Capital: A Critique of Political Economy*. Volume III. Harmondsworth: Penguin Books.
Roemer, J. 1982. *A General Theory of Exploitation and Class*. Cambridge, MA: Harvard University Press.
Roemer, J. 1986. *Value, Exploitation and Class*. London: Harwood Academic Publishers.
Roemer, J. 1996. *Egalitarian Perspectives*. Cambridge: Cambridge University Press.
Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica* 50: 97–109.
Rubinstein, A. and A. Wolinsky. 1985. Equilibrium in a market with sequential bargaining. *Econometrica* 53: 1133–1150.
Samwick, A. 1998. Discount rate heterogeneity and social security reform. *Journal of Development Economics* 57: 117–146.
Scaramozzino, P. 1991. Bargaining with outside options: wages and employment in UK manufacturing. *Economic Journal* 101: 331–342.
Skillman, G. 2016. Wealth inequality and economic power: a sequential bargaining analysis. *Metroeconomica* 67: 291–312.
Vrousalis, N. 2013. Exploitation, vulnerability, and social domination. *Philosophy and Public Affairs* 41: 131–157.

APPENDIX: Proofs of Propositions

Proposition 1. *There is a unique subgame-perfect equilibrium in which bargaining concludes immediately and the expected payoff to an agent of type j is* $\Pi_j^* = (V - A_{-j} + A_j)/2,\ j = K, L.$

***Proof*:** Consider first the proper subgame that begins after the agent making the initial offer has been randomly chosen. Call this agent j and the initial respondent $-j$, noting that j can represent either K or L. Let the time period τ in which the initial offer is made be normalized to 0. Assume that equilibrium exists for this subgame, and let m represent the infimum of possible (expected) payoffs received by j in any equilibrium. Because time-discount bargaining costs are sunk, the pattern of available moves across time is stationary, and the production surplus is not created and shared until all bargaining has concluded, this subgame has a recursive structure that is re-encountered at every even value of t, in which j makes an offer by supposition. Thus, agent j can expect to receive at least m at $t = 2$.

At $t = 1$, therefore, upon rejecting the current offer from $-j$, agent j can expect to receive an immediate payoff of a_j and a payoff of at least m in the following period, with a combined discounted payoff value to j of at least $a_j + \delta m$. By the condition of subgame perfection, $-j$ must offer j at least this amount in order to get j to accept $-j$'s offer in period 1. Consequently, the most that $-j$ can expect to receive in this period is $V - a_j - \delta m$.

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

Thus, the most that agent -*j* can expect to receive at $t = 0$ upon rejecting *j*'s current offer in this period is $a_{-j} + \delta \cdot (V - a_j - \delta m)$. This is therefore the most that *j* must offer -*j* in order for –*j* to accept the offer, under the requirement of subgame perfection. In turn, then, the least that *j* can expect at $t = 0$ is $V - a_{-j} - \delta \cdot (V - a_j - \delta m) = (1 - \delta) \cdot V + \delta a_j - a_{-j} + \delta^2 m$. But given the recursive structure of the bargaining game, this expression is equal to *m*. Solving this equation for *m* and doing some additional algebraic manipulation yields the implication that if an equilibrium exists, the *minimum* payoff that *j* can expect in any equilibrium is $\hat{m} = [(V - A_{-j}) + \delta A_j]/(1 + \delta)$.

Now let *m* represent the *supremum* of possible (expected) payoffs received by *j* in any equilibrium and repeat the preceding backward induction argument from period 2, replacing 'the least' with 'the most' and vice-versa. This yields the inference that $\hat{m} = [(V - A_{-j}) + \delta A_j]/(1 + \delta)$ is also the *maximum* payoff that agent *j* can expect to receive in any equilibrium, so that if an equilibrium exists for this subgame, it yields a unique distribution. It is easy to show that this distribution can be supported by subgame-perfect strategies in which *j* offers $V - \hat{m} = [\delta \cdot (V - A_j) + A_{-j}]/(1 + \delta)$ whenever advancing an offer and rejects any offer less than $a_j + \delta \hat{m} = [\delta(V - A_{-j}) + A_j]/(1 + \delta)$. Thus, a unique subgame-perfect equilibrium exists for this subgame with payoffs $\pi_j^{j*} = [(V - A_{-j}) + \delta A_j]/(1 + \delta)$ and $\pi_{-j}^{j*} = V - \pi_j^{j*}$ for $j = K, L$.

Since the two subgames are chosen randomly with equal probability, the corresponding expected payoffs for the overall game Γ are $\Pi_j^* = (\pi_j^K * + \pi_j^L *)/2 = [(V - A_{-j}) + A_j/2, \;\; j = K, L$. ■

Proposition 2. *Agent L is economically vulnerable to agent K if and only if* $[\delta(V - A_K) + A_L]/(1 + \delta) < A^f$. *In that case, L is economically exploited by K.*

***Proof*:** Consider the proper subgame in which *K* is randomly selected to make the initial offer. It can be seen from the proof of Proposition 1 that the most that *L* can credibly expect to receive upon rejecting *K*'s initial offer is $a_L + \delta \cdot (V - a_K - \delta \hat{m}) = [\delta(V - A_K) + A_L]/(1 + \delta)$. Thus, if A^f exceeds this value, *L* is economically vulnerable by definition. This establishes the 'if' portion of the proposition. To establish the 'only if' claim, suppose instead that $[\delta(V - A_K) + A_L]/(1 + \delta) \geq A^f$. Then in the subgame under consideration, *L* can credibly secure a payoff sufficient to avoid economic vulnerability. For the proper subgame in which *L* makes the initial offer, it can be seen from the proof of Proposition 1 that the most *K* can expect from rejecting *L*'s offer and making a subsequent counter-offer is $a_K + \delta \cdot (V - a_L - \delta \hat{m}) = [\delta(V - A_L) + A_K]/(1 + \delta)$, so *L* can credibly secure the residual of *V* net of this value, equal to $[(V - A_K) + \delta A_L]/(1 + \delta) > [\delta(V - A_K) + A_L]/(1 + \delta)$. Thus, *L* is not economically vulnerable in either proper subgame. ■

Proposition 3. *The bargaining game with exit options has a unique subgame-perfect equilibrium in which initial offers are immediately accepted and yield the respective payoffs* $(\Pi_K^*, \; \Pi_L^* = V - \Pi_K^*)$, *such that*

(E1) $\Pi_i^* = [V - A_{-i} + A_i]/2$ if $W_i \leq [\delta(V - A_{-i}) + A_i], \;\; i = K, L$;
(E2) $\Pi_i^* = [V - W_{-i} + W_i]/2$ if $W_i > [\delta(V - A_{-i}) + A_i], \;\; i = K, L$; *and*
(E3) $\Pi_j^* = [(1 - \delta)(V - A_{-j}) + (1 + \delta)W_j]/2$ *if* $W_j > [\delta(V - A_{-j}) + A_j]/(1 + \delta)$ *and* $W_{-j} \leq [\delta(V - A_j) + A_{-j}]/(1 + \delta)$, $j = K$ *or* L, *but not both.*

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

Proof of Proposition 3: Let W_j denote the expected payoff to agent j upon exiting the existing bargaining relationship, $j = K \text{ or } L$, and follow the same initial steps as in the proof to Proposition 1 up to the step in which agent j's options at $t = 1$ are considered. At this point, agent j can elect to either reject the standing offer and make a subsequent counteroffer, as before, or exit the relationship entirely, and thus can ensure a payoff of *at least* $\max\{W_j, a_j + \delta m\}$. Player $-j$ can in turn expect to receive *at most* $V - \max\{W_j, a_j + \delta m\} = \min\{ V - W_j,\ V - a_j - \delta m\}$if period 1 is reached. It follows immediately that $-j$ can thus expect to receive *at most* $\max\{W_{-j},\ a_{-j} + \delta \min\{ V - W_j,\ V - a_j - \delta m\}\}$ at $t = 0$, and thus j can expect to receive *at least* $V - \max\{W_{-j},\ a_{-j} + \delta \min\{ V - W_j,\ V - a_j - \delta m\}\}=$ $\min\{V - W_{-j},\ (1 - \delta)V - a_{-j} + \max[\delta W_j,\ \delta a_j + \delta^2 m]\}$ at $t = 0$, which therefore equals m.

Now let m denote the *supremum* of possible (expected) payoffs received by j in any equilibrium and repeat the foregoing argument, replacing *at least* with *at most* and vice-versa. As in the proof for Proposition 1, this yields the same value for m, implying that if an equilibrium exists for this subgame, it yields a unique pair of payoffs. It is then straightforward to show that these payoffs can be supported by a pair of subgame-perfect strategies in which player j proposes to receive m whenever making an offer, and accepts any counter offer yielding at least $z = \max\{W_j,\ \delta_j m\}$; correspondingly, player $-j$ proposes z for j whenever making a counteroffer, and accepts any offer of at least $V - m$. These strategies are subgame-perfect and ensure that bargaining concludes immediately in equilibrium.

There are thus three possible equilibrium scenarios for the subgame in which player j makes the initial offer, depending on the pattern of inequalities among the three terms in the expression for m. For example, $m = (1 - \delta)V - a_{-j} + \delta a_j + \delta^2 m$ if $V - W_{-j} \geq (1 - \delta)V - a_{-j} + \delta a_j + \delta^2 m \geq (1 - \delta)V - a_{-j} + W_j$, implying in turn that $m = [(V - A_{-j}) + \delta A_j]/(1 + \delta) = \pi_j^{j*}$. This is the share the initial offerer proposes for him- or her*self* given that $-j$ cannot credibly threaten to exit in response to an unfavorable offer. Correspondingly, $V - m$, the share received by player $-j$, equals $[\delta(V - A_j) + A_{-j}]/(1 + \delta) = \pi_{-j}^{j*}$. Back substitution of the value of m into the inequality conditions establishes that this equilibrium case occurs if and only if $[(V - A_{-i}) + \delta A_i]/(1 + \delta) \geq W_i\ \forall i$, which is the condition for scenario (E1) of the proposition, which occurs when neither player's outside payoff is sufficiently high for the threat of exit to induce a higher payoff. Since this outcome occurs no matter which agent makes the initial offer, and agents have equal probabilities of making the initial offer, the expected payoff of each player in equilibrium scenario (E1) is $\Pi_i^* = [\pi_i^{i*} + \pi_i^{-i*}]/2 = [V - A_{-i} + A_i]/2,\ i = K, L$, as asserted.

The conditions for the remaining two scenarios are established in similar fashion by considering the implications of the other possible combinations of inequalities. The key difference of these latter cases from the one examined above is that at least one player has a sufficiently high outside payoff to make exit a credible threat for that player in response to an unfavorable offer from the other. ■

Proposition 4. *Agent L is economically vulnerable to agent K if and only if* $\max\{W_L, \min[\delta(V - W_K) + A_L,\ (\delta(V - A_K) + A_L)/(1 + \delta)]\} < A^f$. *In that case, L is economically exploited by K.*

https://doi.org/10.1017/S0266267118000081 Published online by Cambridge University Press

Proof: Follow the same steps as in the proof to Proposition 2, except that in the bargaining game with exit options, $m = \min\{V - W_{-j},\ (1-\delta)V - a_{-j} + \max[\delta W_j,\ \delta a_j + \delta^2 m]\}$, as shown in the proof to Proposition 3. This implies that the most L can expect to receive in the equilibrium of the subgame in which K makes the initial offer is $V - m = \max\{W_L, \min[\delta(V - W_K) + A_L,\ (\delta(V - (1-\delta)A_K) + (1-\delta)A_L) - \delta^2 m]\}$. The sufficiency claim of the proposition then follows from solving for the three possible equilibrium values of m, comparing the resulting values of V-m to A^f, and noting that L is economically vulnerable by definition if $V - m < A^f$. To establish the necessity claim of the proposition, note that if $V - m \geq A^f$ in the equilibrium of the subgame for which K makes the initial offer, then $m \geq A^f$ in the equilibrium of the subgame for which L makes the first offer, due to the first-mover advantage. ■

Proposition 5. *Suppose that the matching function is semi-Walrasian and* $\forall t\ N_{it} = N_i^s > 0,\ i = K, L$. *Then in the limit as* $\varepsilon \to 0$, *semi-stationary equilibrium yields payoffs* $(\Pi_K^*,\ \Pi_L^* = V - \Pi_K^*)$ *such that* $\Pi_j^* \begin{cases} = V - A_{-j} \ if\ N_j^s < N_{-j}^s \\ \in [A_j, V - A_{-j}]\ if\ N_j^s = N_{-j}^s \end{cases},\ j = K\ or\ L.$

Proof: In the case of semi-stationary equilibrium such that $N_j^s < N_{-j}^s$ for a particular agent $j = K$ or L, $q_j = (1-\varepsilon)$ given the semi-Walrasian matching function, implying in turn that $q_j \to 1$ in the limit as ε approaches zero. In contrast, the same property of the matching function implies that $q_{-j} \leq N_j^s / N_{-j}^s < 1$ for all positive values of $\varepsilon \in [0, 1)$ and is thus bounded strictly below 1. Correspondingly, the steady-state value of agent j's expected payoff from exit becomes $W_j^s = [(1-\varepsilon)\Pi_j^s + \varepsilon(1-\delta)A_j]/(1-\delta\varepsilon)$, which approaches Π_j^s in the limit as $\varepsilon \to 0$.

To see that this limit case is inconsistent with equilibrium scenarios (E1) and (E2), assume otherwise. This cannot occur in equilibrium scenario (E1), since in this case $\Pi_j^s = [V - A_{-j} + A_j]/2 \to W_j^s$ as $\varepsilon \to 0$, contradicting the requirement that $W_j^s \leq [\delta(V - A_{-j}) + A_j]/(1+\delta)$ for scenario (E1) to occur. In equilibrium scenario (E2), $\Pi_j^s = [V - W_{-j}^s + W_j^s]/2 \to V - W_{-j}^s$ as $\varepsilon \to 0$, implying in turn that $\Pi_{-j}^s \to W_{-j}^s$ in the same limit. But then $W_{-j}^s \to [q_{-j}^s W_{-j}^s + (1 - q_{-j}^s)(1-\delta)A_{-j}]/(1 - \delta(1 - q_{-j}^s)) = A_{-j}$ as ε approaches zero, which contradicts the condition $W_{-j}^s \geq [\delta(V - A_{-j}) + A_j]/(1+\delta)$ required for scenario (E2) to occur.

This leaves equilibrium scenario (E3), which implies that $\Pi_j^s = [(1-\delta)(V - A_{-j}) + (1+\delta)W_j^s]/2 \to V - A_{-j}$ as $\varepsilon \to 0$, implying in turn that Π_{-j}^* approaches A_{-j} as ε approaches zero. It is readily verified that both inequality requirements for scenario (E3) are satisfied for values of ε sufficiently close to zero.

For the case that $N_j^s = N_{-j}^s$, the semi-Walrasian property of the matching function implies that $q_i^s = (1-\varepsilon)$, $i = K, L$, implying in turn that for all agents, $W_i^s \to \Pi_i^s$ as $\varepsilon \to 0$. As before, this limit case is inconsistent with equilibrium scenario (E1), but is compatible with scenarios (E2) or (E3) for values of Π_j^s in the indicated range, given $\Pi_j^s + \Pi_{-j}^s = V$. ■.

Proposition 6. *In the limit as* $\varepsilon \to 0$ *of semi-stationary equilibrium under the conditions of Proposition 5, agents of type L are exploited if* $N_L^s > N_K^s$, *but are not* *exploited if* $N_L^s < N_K^s$*. If* $N_L^s = N_K^s$*, there are both exploitative and non-exploitative equilibria.*

Proof**:** In the case that $N_L^s > N_K^s$, the limiting semi-stationary equilibrium is such that agents of type L can credibly secure no more than A_L no matter who makes the initial offer. Since $A_L < A^f$, agents of type L are thus economically vulnerable, and are exploited in equilibrium since the payoff to agents of type K is inversely related to A_L. In the case that $N_L^s < N_K^s$, agents of type L can always credibly secure $V - A_K > A^f$ no matter who makes the initial offer, and are thus never economically vulnerable or exploited. Finally, if $N_L^s = N_K^s$, both of the above equilibria are attainable, and thus there are both exploitative and non-exploitative equilibria. ■

BIOGRAPHICAL INFORMATION

Gilbert L. Skillman is Professor of Economics at Wesleyan University. He is the author, with Joyce P. Jacobsen, of *Labor Markets and Employment Relationships* (Blackwell Publishing, 2004). His current research involves the application of game-theoretic analysis to the study of topics in political economy and labour economics.