The creation of jobs
Michael Joffe1 [Imperial College, London, UK]

Abstract
This paper investigates the source of jobs in the modern economy, excluding the state, non-profit and financial sectors. The approach is centered on the firm and its profitability, and in particular, proposes a real investment led economy perspective. Rather than assuming that productive systems already exist, as is frequently done, it examines how they are established and renewed, and also how they may cease to exist. Investment is seen as the strategic decision process that establishes the number and type of jobs for the medium term, and their approximate wage level. Thus, at the time of the job creation decision, it is not clear who the potential workers are – and particularly when technological change and off-shoring are involved, even what population they will be drawn from. Investment often involves a business plan: a joint decision about the product, location, technology, scale of production, etc, as well as employment. In addition, tactical decisions are taken during the course of production. Both types of decision take account of the economic environment, and involve Knightian uncertainty. Unemployment exists when fewer jobs are created than there are people who would like to fill them, because there are too few investment opportunities that are perceived as being potentially profitable. In the investment decision, the whole package needs to be coherent, and offering lower wages may not be enough to make an investment idea potentially profitable. This explains why non-frictional unemployment can occur, and persist, especially when perceived investment opportunities are few. It involves asymmetry: a shortage of workers is reflected in an increase in the wage level, whereas a shortage of jobs is manifest in terms of unemployment – quantity not price. The real investment led economy perspective is also able to account for employment changes in six major types of scenario: a new company, major technological change, relocation, plant closure, regional decline and a major depression. Existing theories struggle to explain these phenomena. In addition, this perspective naturally addresses several puzzles in standard labor economics.

JEL codes J21, J23, J31, J60, E24, L21

Key words firm-centered approach, real investment led economy, unemployment, technological change, offshoring, labor market puzzles, wage flexibility puzzle, unemployment volatility puzzle, employer size-wage effect (ESWE)

1. A long term dynamic view of the labor market

A fundamental issue in economics is the source of jobs in a modern capitalist economy: where do they come from? Who creates them? When? Why? How? A comprehensive theory of the labor market needs to be able to answer these questions in a satisfactory manner.

In addition to the context of business as usual, such a theory must engage with the major situations that affect employment, unemployment and wage levels. As well as accurately describing routine operation, it must be able to give a good account of the implications for employment and wages under conditions that involve alteration of the trajectory of a firm or an economy. This means the ability to encompass the following scenarios, four of which represent the key strategic decisions that firms take, the other two being broader conditions that impact firms and their profitability:

---

1 I would like to thank Victoria Chick and Jon Fjeld for helpful discussions and comments on a previous draft.
(a) setting up a new company;
(b) introducing a major technological change;
(c) relocation of production;
(d) a plant closure;
(e) an economically depressed region;
(f) a major depression.

This paper introduces a new perspective on employment, unemployment and wage setting that can naturally meet this challenge. It is centered on the firm and its decision making, rather than on the relationship between firms and workers, because it is the firm that takes the major initiatives that affect the location, quantity, type, quality and remuneration of jobs. The central consideration for a firm is its profitability, as survival is impossible for long when making a loss. The primary focus here is on the investment process undertaken by firms from time to time, including the setting up of new firms. This viewpoint could be called the real investment led economy perspective. It is derived from a description of the observed behavior of firms, especially the time order of key decisions and events, and is evidence-based in this sense.

The term “real” is included to specify investment in the real economy, rather than financial or real estate investments – the focus is on productive firms. The description applies not only within manufacturing, but also to private sector non-financial services. I am neglecting the public sector and non-profit firms in this paper.

The major traditions in labor economics, based e.g. on the neoclassical theory of the firm and on the search-and-matching approach, are well established. But they share the feature that a production system is assumed to be already in existence. In the case of neoclassical theory, the focus is on the decision to employ an extra worker, or one fewer, at the margin, given a production set-up that is already a going concern. In the search-and-matching perspective, current employment positions and prospective workers are brought together, which requires the assumption that the jobs already exist, and the workers are appropriate for them. This implies a limitation in both cases, which is particularly clear in a situation where a firm may decide between different options for locating the new production, or where there are substantially different possible technologies that have major implications for the number and type of workers that would be required. Under those conditions, it is not even obvious who the prospective workers are, until the location and/or technology decision is taken.

The assumption that employment and wages can be validly analyzed in the context of a productive setup that already exists could be justified for a relatively simple economy, for example medieval and early modern England, e.g. where one is discussing the hiring of different types of building workers (Clark, 2005). However, in the modern economy most types of occupation have been introduced by capitalist firms at some stage, often combined with new technology and/or new products. This was true of employees in the early 19th century cotton mills of Lancashire, of Henry Ford’s production line workers, and remains true today, e.g. of workers in Shenzhen assembling iPhones. The modern production system continually creates new forms of employment, as well as new technology and new products. Even for the large majority of capitalist firms that are not innovative in this respect, and are merely treading the already-established paths, their investments are the primary source of the private sector jobs created in the real economy. Accordingly, the real investment led economy view traces the source of employment opportunities to the initiatives taken by capitalist firms, even when they are relatively routine ones.
The real investment led economy perspective maintains that investment is endogenous to the system, an integral part of it. Although investment may appear to be an autonomous decision, it responds to external stimuli in the economic environment – the institutional and technological environment, and especially the projected demand for the product. This latter provides a link with aggregate demand at the macro level. A hallmark of this viewpoint is that it is dynamic: it is centrally concerned with how economies change over time. The analysis focuses as much on how productive systems come into existence, and how they change, as on more short-term questions about how they operate.

The real investment led economy perspective is the positive side of the more general firm-centered view. For relatively strong firms in a favorable situation, the focus is on investment. The negative side applies to firms in a weak position, resulting from their lack of capacity and/or the lack of demand for their product(s). It similarly emphasizes the central causal importance of the firm and its decisions, the primary aim being to make profits or at least to avoid losses.

2. An outline of the real investment led economy perspective

The core idea in the real investment led economy perspective is that the real economy has been created over time by the initiatives taken by firms. This occurs in the process of investment, either by existing firms or by start-up companies. Investment is the occasion on which those who control the firm make the key strategic decisions that set the pattern for its future operation. This is a single many-faceted decision, nowadays often specified in a business plan. At the time that the plan is developed, it is based on the best information that can reasonably be obtained, and can therefore be assumed to be optimal in this sense. The centrality of investment is a characteristic of the modern “capitalist” economy.

The decisions include the nature of the product(s), location, technology, broad scale of production, and types of labor – the numbers required of each skill type and an expectation of their approximate wage levels. One of the primary concerns is recruitment and retention of suitable workers, in the right numbers, at the right wage, in the proposed location. A business plan also analyses the availability of finance, both internal and external, and estimates the potential sales quantity and price. The wage and price decisions are made in comparison with what others are doing, including the going wage for each type of employment. Ideally, it also takes into account how this is likely to evolve.

Crucially, all these components need to fit together in a way that promises success – the sums need to add up, and the strategy needs to be coherent. In particular, decisions involving employment type and quality cannot be separated from those on technology. For example, there is evidence that investment in information technology involves also the adjustment of work practices as well as the rethinking of product offerings, a process that requires experimentation (Bresnahan et al., 2002). And various components of workers’ skill are related to firms’ technological inputs (Abowd et al., 2007).

The concept of investment is broader than the standard image of purchasing equipment for the purpose of manufacturing. It includes starting a hotel or restaurant, or setting up a homecare agency, or an airline.
Not all investment is aimed at increasing (or transforming) productive capacity. Auxiliary investment, for example a marketing initiative aimed at increasing sales of a current product line, enhances the value of an existing investment in productive capacity. This can also affect employment.

An investment therefore involves far more than just buying some machines. This contrasts with the thought experiment frequently found in textbooks, in which a firm decides separately on a fixed and a variable factor of production, usually capital and labor, respectively.

All aspects of the investment are set in accordance with the requirement of future profitability. This will clearly depend on expected future market conditions, relating both to the inputs to production and to the product, that can only be estimated rather roughly at the time that the investment is being planned due to the presence of Knightian uncertainty. This central idea is based on the non-controversial statement that firms invest so as to maximize expected future profit, taking into account the economic environment.2

The economic environment includes the types and quality of workers available in particular candidate locations, and the likely cost of employing the required number for the proposed scale of production. The major employment decision therefore rests on the potential productivity of the available workers compared with the wage level required to attract them. The going wage must be taken as given, taking account of the possibilities of offering extra in order to enhance worker quality, motivation and retention, and of augmenting existing human capital by training programs – seen in the broader context that includes location and technology. In this way, firms create employment positions, or in the terminology of Acemoglu and Autor (2011), occupations which are bundles of tasks.

3. The real investment process

Firms invest when the expectation of future gains exceeds the expectation of the necessary costs by an amount that generates a satisfactory return on capital. Costs here include fixed costs as well as variable costs, and taxes. Investment is a strategic decision that sets the firm’s future direction. It includes the following:

1. Employment depends on the firm’s investment decision – this neglects the cases in which a new position is created in response to an initiative taken by a potential worker, which may sometimes occur (especially in the case of very highly-skilled/specialized people), but is probably unusual.

2. Numbers of workers of each skill type are similarly part of the firm’s investment decision. This is a joint decision with the technology decision, e.g. how many machines are to be purchased.

3. The approximate wage level for each type of worker is also part of the firm’s investment decision, taking account of the going wage, plus a possible increment to improve recruitment, motivation and retention if that can be afforded. This may involve a joint

2 Actually, maximization is a convenient assumption but is not an essential feature; this perspective could be compatible with e.g. a satisficing view of firms’ behavior.
decision with location, especially for large firms that are able to (re)locate abroad, to a country with a beneficial ratio of worker productivity to wage level.

4. Investment depends on the availability of finance, including retained profits as well as possible external funding. The wage level is often specified as a wage range, i.e. it is approximate ex ante, giving scope for negotiation with workers once they are identified, and/or with their representatives, which corresponds to the wage-bargaining situation in the literature; on the other hand, when a precise wage is specified as a condition of employment, this corresponds to wage posting (Hall and Krueger 2008). In this perspective, they are not regarded as polar opposites — rather, the difference is the comparatively minor distinction between approximate and precise specification of the wage level, respectively. Firms also make the decision on how to structure the employment relationship in such a way as to try and best motivate employees (Oyer and Schaefer, 2011).

The ability to make a potentially profitable investment depends on the firm’s capacities, including its managerial capacity (Penrose, 1959), and its strength relative to others in the context of the power struggle between competing firms (Joffe, 2011). With relatively weak firms, the situation can arise that a firm may go for years without investing, i.e. there is stasis and possible decline. This could be for lack of funds and/or possible investment opportunities that promise improvement, or it could just be a question of inadequate or complacent management. The consequence for workers is likely to be stagnation in wages, and ultimately insecurity of employment if the firm then struggles to survive. Similarly, in a sector that is disrupted by the introduction of substantially lower costs or new/improved products, incumbent firms need to react, e.g. by trying to adopt the new practices, or by scaling back production and/or restructuring. These conditions affect firms that are in a weak position relative to competitors – they are on the back foot – and are therefore less able to take initiatives that are potentially profitable. Whilst strictly speaking not within the real investment led economy viewpoint, it is still best seen in a firm-centered perspective, with a primary focus on the degree of profitability.

4. Decision making during the course of production

In addition to the strategic decision making involved in investment, more day-to-day tactical decisions are subsequently taken in the course of production. These include recruiting workers to replace those who have decided to leave; dealing with disciplinary issues and conflict; maintenance and repair of equipment and premises; and responding to current market conditions, e.g. by reducing the price of items that are not selling.

The production-related component includes the following items that involve the labor market:

1. The exact wage level may be subject to negotiation with workers once contact is made with them. Generally speaking, workers with higher-level or scarce skills tend more often to be able to negotiate, because they are in a better bargaining position than easily-replaceable low-skilled workers (Hall and Krueger, 2008).

3 The topic of the relationship between competitor firms, and in particular their relative strength, is beyond the scope of the current paper – for a discussion, see Joffe [2011], especially section 3.1.
2. The wage level may be subject to negotiation in the context of new matches during the life course of an investment, i.e. workers recruited at some later date, for example to replace departed employees; this is likely to be related to market conditions such as the unemployment rate (Pissarides, 2009). Renegotiation with existing workers may also occur, e.g. in an annual review.

3. The number of hours worked per worker is similarly subject to negotiation/renegotiation and to current market conditions.

Decisions that are not directly related to the labor market similarly have both strategic and tactical components, e.g. the quantity and price of the output, and the sourcing of raw materials.

Any satisfactory account of the process of job creation needs to combine the strategic decisions with the tactical modifications that occur during the lifetime of the investment. These latter typically involve workers (individually and/or via their representatives) as well as firms, and are amenable to a more conventional approach, e.g. using the canonical search and matching framework (Pissarides, 2000).

From the firm’s viewpoint, production decisions are taken in the light of the economic environment, just as investment ones are. For example, a firm may decide to increase wages in order to poach or retain workers, in relation to the actions of other firms (Moscarini and Postel-Vinay, 2008). The "economic environment" here includes the financial position of the firm – what capital it can raise in addition to its own retained funds – and market conditions, i.e. what it will be able to sell, as well as supply-chain factors, and what competitors are doing.

Both investment and production decisions involve the passage of time: firms make their decisions on the basis of their expectations of what is likely to be profitable, and workers then respond. They also both involve Knightian uncertainty. However, they differ greatly both in their time horizon and in their degree of uncertainty.

Tactical decisions in the course of production tend to have a time horizon of days, weeks or months, and the degree of uncertainty in the outcome is moderate. Strategic investment decisions are accompanied by a short-term boost to employment in the sectors that receive the expenditure, e.g. equipment manufacturing or construction. Its longer-term impact on employment is uncertain in two distinct ways. From the firm’s own viewpoint, an expectation is formed during the investment process, and then the practical realization depends on the degree of success of the investment in generating new economic activity that finds favor in the market in the ensuing years – and the degree of success is typically highly uncertain at the time of the investment decision. From the viewpoint of the wider economy, the impact on employment is ambiguous: with investment in labor-saving technology, it may decline even if the investment is successful in the firm’s terms.

5. Why unemployment exists

An implication of this view is that the overall employment available in an economy is the total of that provided by all the investments. In this perspective, unemployment is seen as the gap between the positions created at any one time by firms and the number of people who are available for work, stratified by skill type and location. Thus, unemployment exists because
insufficient investment opportunities are perceived to exist that promise to be profitable with the available combination of workers’ abilities and the necessary wages, as well as other conditions such as the prospective market and the costs of raw materials, fuel and taxes. Its level depends on the perceived scope for profitable investments available to firms within that economy.

Some long-term correspondence of employment and working-age population exists, due to a process of adjustment. If the economic environment includes a tight labor market with high wages, this could be a deterrent to investment, for some firms at least. The converse is not necessarily true, because even with low wages, potentially profitable investments may be limited. The combination of available firm-based resources, including managerial talent and available technology, with the potential demand for the product may not add up to an investment that promises to pay off, even with low wages. The adjustment process is thus asymmetric. (It is also slow, because investments are infrequent, and designed to last for a long time.) The asymmetry means that labor scarcity is typically manifest as a rise in wage level, whereas labor abundance remains a matter of quantity rather than price.

The corollary is that some unemployed workers are likely to be unemployed because no additional investments are seen as possible even with lower wage levels, and even if potential workers would be willing to accept lower wages. As a result, in the aggregate, there are too few jobs to go around. A related question concerns the distribution of jobs: one might think that the workers who happen to be unemployed would be able to exchange their positions with already-employed workers by offering to work for a lower wage. But wage undercutting seldom occurs. Why?

There are three ways that wage undercutting could operate: (i) existing workers could agree to take a pay cut, (ii) unemployed workers could seek to undercut the existing ones in an existing firm, and (iii) a new firm could enter the market, with lower-paid workers.

(i) is unusual, but has been observed e.g. when a firm is in serious trouble, and the alternative is closure or large-scale redundancies.

(ii) could occur (a) because the firm’s previous investment decision was suboptimal, i.e. the jobs that were newly created or maintained could have been offered at a lower wage. Such an error in decision making may occur from time to time, but in general it is probably safe to assume that firms’ decisions at the time of investment are close to optimal, given their managerial capacity and the uncertain nature of information about future economic conditions. Alternatively, (b) there could be a major change after the investment was made, e.g. a shock that resulted in the availability of an unforeseen pool of potential workers with the right skills, in the right location. In such a case, this shock would be the principal cause of the possible undercutting situation, rather than the mere existence of currently unemployed potential workers who are willing to accept lower pay.

(iii) does occur in some industries, as new firms enter the market with a new business model, a more efficient technology, or a more cost-effective location (see Scenarios, below). In such cases, decision making on wages is typically integrated with decision making on broader work practices, as well as with other factors such as technology and location.
Another implication is that the behavior of individual potential workers does not impact the number or type of jobs. An unemployed person who is particularly diligent in job search is more likely to achieve employment, but this is merely at the expense of someone else.

In summary, firms are seen as proactive, forward-looking agents that take the initiative. In contrast, (potential) workers are seen as reactive, in the sense that they react to the employment positions created by firms – acknowledging that the decisions which create these positions are arrived at by taking into account the existence and characteristics (number, skill type, going wage, etc) of the potential workers. These could currently be unemployed, employed elsewhere, self-employed or inactive. Firms propose; workers react.

This perspective is supported by evidence that workers tend to be backward looking in their reservation wages, possibly driven by perceptions of fairness and/or backward-looking reference points (Akerlof and Yellen, 1990; Falk et al., 2006; Della Vigna et al., 2014; Koenig et al., 2016). The past orientation of workers, together with the setting of approximate wage levels by firms as a strategic decision at the time of investment, intended to last for its whole duration, imply a substantial degree of inertia in wage levels.

6. The broader context

The real investment led economy viewpoint only applies to a real economy dominated by capitalist firms, in the sense of firms that are able to buy in all their inputs, including labor. It is well recognized that unemployment in this sense is a relatively recent phenomenon, since the industrial revolution. Outside the capitalist context, e.g. in peasant agriculture, the equivalent is underemployment, which typically occurs when the working-age population is large compared to key resources such as land, rather than to the size of workforce implied by the sum total of investment.

Non-capitalist firms survive in the modern context as self-employed individuals. Confusingly, they are referred to as “entrepreneurs”, leading to widespread conflation with the idea of innovative entrants who bring about creative destruction. Entrepreneurs in the former sense are generally observed to have lower productivity than capitalist firms (e.g. GEM, 2017). It is also a myth that most innovation is due to entrepreneurs in the sense of outsiders (Hsieh and Klenow, 2017).

One implication of the real investment led economy viewpoint is the possibility that firms may have funds available to invest from retained profits, but lack suitable investment possibilities for future production. The possible consequences are that they will invest in property or the financial sector instead of in the real economy (“financialization”), and that if this occurs on a widespread scale, secular stagnation of the real economy could result.

Another implication concerns the nature of investments – they are not all equal in their consequences for the jobs they create. Some investments provide employment with high productivity, good pay, and a degree of job security. Others only create insecure, low-productivity jobs, where the main aim is to minimize the wage bill.

Substantial changes in the labor market of many rich countries have occurred in recent decades. A firm may decide to classify its workers as self employed in order to avoid legal requirements relating to sick pay, taxation, etc., to reduce costs. This may be accompanied by
flexibility in the number of work hours offered to the workers, e.g. zero-hours contracts, a way of fine-tuning the expenditure on labor to the current demand. This situation is created in the course of investment, e.g. in setting up a minicab firm with its office etc., that is geared up for engaging drivers as and when they are needed. There is actually nothing new in this strategy, which was present in the early twentieth century, e.g. in dock work in Britain.

Its most recent form involves the use of software – also a type of investment – such as Uber or Deliveroo. In some of these cases, the authority relationship of the capitalist firm is retained, e.g. in Uber’s control over what drivers do, its use of ratings, the requirement to take a certain proportion of clients, etc. However, the situation is somewhat different from the “traditional” form of employment that had (approximately) guaranteed hours, because the variability in the number of available work hours raises the possibility of underemployment.

The type of employment created also has implications for socioeconomic mobility at the societal level. In particular, at a stage in the development of an economy where managerial and professional jobs are created on a hitherto unprecedented scale, the possibilities for upward mobility are transformed. In the real investment led economy perspective, large-scale upward social mobility is a matter of the available positions – rather than of the attributes of the upwardly mobile individuals, the nature of their education, etc. Clearly, both aspects are important in practice.

7. The relationship of the real investment led economy perspective to existing views

This perspective contrasts with standard neoclassical theory in several respects. That theory puts forward models relating to the decision making of (potential) workers, and of firms – respectively the supply of and the demand for labor. Workers choose whether or not to accept employment, based on a comparison of the offered wage with their reservation wage. Firms’ decision making is seen as a comparison between employing one more or one fewer worker with the difference this would make to production – respectively marginal cost and marginal benefit – given that the firm already exists, and has an established production system with premises, equipment, etc. Neoclassical theory implies that the forces of demand and supply rapidly bring about an equilibrium in which there is neither excess demand for labor, nor excess supply.

The real investment led perspective holds that it is true that a process of adjustment does occur (see Section 5), but the timescale is far slower than neoclassical theory would imply. It involves investment decision making that depends on the economic environment, including likely future demand – the micro (or meso) analog of Keynesian aggregate demand. In addition, the bulk nature of investment implies that the decision making is “lumpy”, rather than the “smooth” process implied by the theory – although very small firms’ decisions may approximate to the notion of adding or removing an individual marginal worker. Finally, it is an open question whether investment is best seen as part of an adjustment process leading towards equilibrium, or as a disequilibrium process of creative destruction – a topic that is beyond the scope of the present paper.

---

4 In Britain, a minicab company is a firm that coordinates the work of self-employed owner drivers. They pay a fee to the company for the services of the call centre, which takes bookings and schedules the work. The firm usually supplies radios as well. Each driver has to wait to be told when a trip is available, and a standard flat-rate charge applies.
In the neoclassical perspective, wages are – or “should be” – flexible, and the labor market “should” clear. The observation that persistent unemployment is quite frequently observed has led to the modification that (nominal) wages are assumed to be sticky downwards. This asymmetry can be compared with the asymmetry in the real investment led economy perspective, which is because some investments do not promise to be profitable even with low wages, so that a shortage of jobs does not lead to a fall in the wage level; in contrast, a shortage of workers does raise wages (section 5).

As its name suggests, the real investment led economy perspective places investment at the center of the economy. A large proportion of economic activity is the result of previous investments by firms. This means that the investment decision, the driving force, is strongly endogenous to this system in a causal sense. This contrasts with the view that attributes labor market phenomena to shocks to e.g. demand or productivity.

The familiar concept of rent attributable to employer-employee matches, divided between the two parties, is represented here by (a) a contribution to firm profitability, and (b) comparison with the worker's outside options, i.e. when unemployed, inactive or self-employed, or the wage in a previous job (for matches following in-job search), as appropriate.

In the literature, some emphasis has been placed on the distinction between new jobs and existing or continuing jobs. The proposed perspective suggests the need to further distinguish between jobs that are newly-created (strategic) and newly-negotiated (tactical – e.g. new matches). It explains why wages in new matches (tactical) respond to the current unemployment level, whereas those in existing jobs do not (Pissarides, 2009; Koenig et al., 2016) – they were already set during an earlier strategic decision.

The real investment led economy view takes labor supply as given. It does not address participation rates, i.e. flows between unemployment and inactivity, or in-job search as contrasted with the job search of unemployed potential workers. For these aspects of the labor market, a complementary approach such as a search-and-matching model is required. This does have consequences for real-economy investment, via the economic environment. For example, if previously inactive workers become available for work (“unemployed”), e.g. as a result of a policy intervention, or if largescale immigration suddenly occurs, firms can take this into account in their investment decision making for the future.

The real investment led economy view has little to say about how jobs come to an end for an individual worker. This is because the starting point in that situation is an employee who is already in a relationship with an employer, implying a much more symmetrical situation: either party can bring the arrangement to an end. The employee can leave for personal reasons or to move to a better job, or s/he can be sacked for low productivity, indiscipline, etc. But it is highly relevant to larger scale job losses, as discussed in the next section.

8. Scenarios

As previously mentioned, any attempt to explain where jobs come from needs to be applicable to a range of scenarios. They are described in this section, each with an example drawn from real-world experience. There is some overlap between them. The subsequent two sections deal with how these are treated in existing theories and in the real investment led economy perspective, respectively.
(a) Setting up a new company

In recent decades, a number of low-cost "no frills" airlines have been set up. These seek to exploit the high costs of traditional airlines, and the opportunities afforded by deregulation in many jurisdictions. Here we are concerned with the investment process, and its consequences for the labor market. To set up an airline, it is necessary to acquire (buy or lease) aircraft, obtain landing slots and check-in facilities at suitable airports, and obtain government approval for safe operation. All types of staff need to be recruited and trained: aircrew, cabin attendants, ground and maintenance staff, etc. Other costs include fuel contracts, advertising and insurance. All these need to be budgeted for, and finance arranged if not already available. Strategic decisions are necessary on schedules and fares, as well as logo and livery, seating types, meal provision, etc. (See e.g. Creaton, 2004 on the setting up of Ryanair; also The Economist, 2011.) Low costs have been achieved using such policies as quick turnaround times to maximize utilization of fixed capital, the use of cheaper airports, and restricting the air fleet to just one type of aircraft so as to minimize the costs of staff training and the required inventories of spare parts.

One result is the creation of the required number of jobs in each category. A strategic decision needs to be made on the wage level that is deemed necessary to attract, motivate and retain staff of the appropriate skills and quality. This is likely to be strongly influenced by a comparison with similar jobs in similar organizations. Other decisions include policies affecting staff morale, including unionization – different airlines have taken very different directions in this respect. The staff-related decisions are intrinsically bound up with all the other aspects, as in the decision on aircraft types which affects both training costs and maintenance costs.

(b) Introducing a major technological change

Modern technology has made it easy for people to arrange their own bookings, for flights, hotel rooms and myriad other purposes. This has led to major labor market changes. In the hotel industry, for example, a whole stratum of middle management has disappeared as the work of booking rooms has been devolved to customers, as well as being hugely simplified. This process continues to expand, with self-service checking in now being introduced, which has further implications for staffing levels (Worgull, 2017). In addition to the loss of hotel management jobs, some employment is created in software-based firms such as booking.com, but on a much smaller scale.

The loss of employment opportunities here is a by-product of investment decisions of the software-based firms, and the way that hotels react to this new situation. The loss of staff positions clearly entails a large cost saving, but there is no corresponding loss of benefit to the hotel. Neither is there necessarily any deterioration of the service: self-booking is plausibly more convenient for customers, especially as it can be done from a website comparing the availability and prices of different options for any particular location, using a rather small number of clicks. Arguably this is less arduous than having to contact individual hotels by telephone in order to find out if there is a vacancy, and then to make a booking. It also has an added advantage in providing an automatic record and confirmation of the booking.
(c) Relocation of production

Firms may decide to transfer production from one location to another. This can take many forms, e.g. the twentieth century move of much manufacturing within the US, from the traditional industrial regions in the north to areas further south that had lower costs (see below). More recently, it has tended to involve an international relocation, offshoring, especially for large firms.

One example is the decision by Dyson, the British domestic appliance manufacturer. In 2002, vacuum cleaner production was moved from England to Malaysia with the transfer of 800 production jobs. The decision was made on the basis that cheaper labor was available in Malaysia that would not involve a proportionate loss of productivity (Gribben, 2003). The result was an increase in profits and a rise in the number of high-skilled jobs (e.g. in engineering) in England, as well as an expansion into the US market (Gow, 2003).

(d) Plant closure

Plant closures are a major reason for job loss, and their occurrence is strongly cyclical (Davis et al 1998). They often cluster: other plants of the same type in the same industry and country frequently close at around the same time. An example is the closure of integrated steel mills in the US in the 1970s and 1980s. Recessions occurred in 1973-75 and 1981-82, greatly reducing the demand for steel, at a time of rising energy prices following the oil crisis. Other underlying causes included competition from East Asia, and the introduction of a new technology, electric-powered minimills, that competed with integrated mills in low grade steel production. Both these factors gradually increased during the 1970s and 1980s, and the recessions – especially the later one – precipitated massive losses (three billion dollars in 1982) and largescale plant closures. Output shrank by more than a third, and 150,000 workers were made redundant (Rowe, 2016). There was “marked uncertainty about the profitability of new capital investment” (Davis et al., 1998).

Plant closure is the mirror image of investment. Firms that are in a strong position – perhaps because of low unit costs, or because they have a successful new product – are able to take initiatives that are potentially profitable. They invest, and one result may well be that their weaker competitors are forced onto the defensive, and have to scale back production, or to try and imitate their more successful rivals. This may involve restructuring. When all responses fail, the result is likely to be plant closure, or even the demise of the firm.

(e) An economically depressed region

Formerly known as the Manufacturing Belt, the Rust Belt is a large area in the US from the Great Lakes to the upper Midwest States. Its proximity to iron ore and Appalachian coalfields, and to the Great Lakes and other transportation infrastructure, propelled it to industrial prosperity in the early twentieth century. However, from the middle of the century, it began to decline due to a combination of factors including relocation of production to the southeastern states, automation that reduced the need for labor, and trade liberalization that encouraged offshoring (Crandall, 1993; High, 2003). Manufacturing employment fell by a third by 1996 and has continued to decline since then. The result has been not only a loss of jobs, but also a fall in median household incomes of approximately a quarter in some cities. In addition, there has been large-scale out-migration from the former major industrial centers – Cleveland, Detroit, Buffalo and Pittsburgh lost about 45% of their population. Other consequences
included declining tax revenues, swelling welfare rolls, a poor standard of education, and social problems such as crime and drugs.

(f) A major depression

The story of the Great Depression of the 1930s is well known. The boom years of the 1920s ended with the 1929 New York stock market crash. Many countries were devastated in the depression that followed. Unemployment rose to over 20% in the US (Garraty, 1986; Hamilton, 1987). Some of the job losses were immediate, due to the failure of businesses, but beyond that, the failure of investment meant that future production, and therefore future employment, was compromised. This is not the place to discuss the various theories on the deeper causes of the Great Depression, but it is clear that the proximate cause of the medium-term fall in employment was largely a shortfall in real-economy investment – whether that was in turn due to inadequacy of aggregate demand or of the money supply, to debt deflation, to pessimistic expectations, or some combination of these and other factors.

9. Existing theories and the scenarios

(i) Standard neoclassical theory

In standard theory, there is symmetry between firms and workers: their choices play an equally important role. Workers’ choices could be relevant in the first three of the above scenarios: they can decide whether or not to join a new firm, including a software-based hotel booking firm, or to take employment in a plant that has relocated to their area. However, this could only occur after the investment decision had been made and announced. The potential for it to occur would affect firms’ decision making, but only indirectly, via their perception of the availability and cost of the workers they require. The decision making of workers is clearly irrelevant in the scenarios of a plant closure and a depressed region. In the Great Depression, the suggestion would have to be that workers voluntarily left their jobs, and refused new ones, because the offered wages had become too low, so that they preferred “leisure”. This conflicts with the evidence on what happened in the 1930s, e.g. on the shortage of vacancies. More broadly, the causes of involuntary layoffs, as in plant closures, are totally different from those of voluntary quits: for example, the former increase in recessions, whereas the latter fall sharply (Akerlof et al., 1988; Davis et al., 1998).

Firms’ decisions in the scenarios are also at variance with neoclassical theory. In the case of a new airline, the textbook notion that the firm decides on the flexible factor, labor, to fill an already-established number of slots (e.g. machines) corresponds poorly with actual decision making: the decisions on aircraft, slots, etc entail a corresponding complement of the various types of staff required – the business plan specifies everything simultaneously.

In the case of the hotel staff, there is no calculation of marginal costs and benefits. The elimination of the need to pay managerial staff is not accompanied by a loss of benefit to the firm (nor by a loss of convenience to the customer), so there cannot be a trade-off. The relocation of Dyson could be said to involve a comparison of marginal costs and marginal

---

5 A vast number of theories have attempted to explain unemployment – see e.g. De Vroey [2004]. I have therefore been highly selective, focusing on those that have been highly influential and those that appear to be the most persuasive.
benefit, but this would encompass premises, equipment and supply chains, not just employment.

With a plant closure, or a wider downturn in space or time, the firm is typically faced with losses that cannot be allowed to continue indefinitely, and economic conditions where investment does not promise to be profitable. It is possible to describe this in marginalist terms, e.g. that the losses correspond to the difference between marginal cost and marginal benefit, but this is at the plant level (lumpy), not a question of an individual marginal worker.

(ii) Efficiency wage theories

Various theories have been proposed, that firms pay a premium “efficiency” wage. This could be to discourage quitting – especially when turnover costs are high (Schlicht, 1978), to raise the potential cost of workers shirking (Shapiro and Stiglitz, 1984), to promote a sense of reciprocity and fairness (Akerlof and Yellen, 1990), and/or to attract higher-quality staff (Schlicht, 2005). There is good evidence that some firms do pay efficiency wages to at least some of their employees (e.g. Raff and Summers, 1987), and receive efficiency gains as a result.

In addition, it has been suggested that this higher wage has a role in creating unemployment: it is above the market-clearing level, and therefore the labor markets fail to clear. This puts the focus on the worker’s decision: a comparison of the wage level and her/his reservation wage, the lowest wage at which s/he will accept the job. However, in the scenarios the appropriate comparison is rather the firm’s decision-making process. The decision to close a plant depends on its loss making (or failure to make adequate profit), which is the difference between revenue and costs; the workers’ reservation wages are irrelevant. Similarly, the absence of investment in a new company, technology or location depends on the going wage, not the reservation wage (which is unknowable to a potential investor), together with other potential costs and the likely revenue – the absence of sufficient promise of profit.

The situation is somewhat different for the last two scenarios, a depressed region or a recession, which involve also the aggregate level. In these circumstances, the higher wage would have some protective effect at the macro level, by raising aggregate demand, thereby reducing the degree of contraction. On balance, a higher wage bill could therefore increase employment.

Thus, the efficiency wage theory may accurately describe much of the behavior of firms – especially the better-off ones, or those that have a relatively low wage bill because they are capital intensive. But it is irrelevant to the perceived need to explain the lack of market clearing – a perception that derives from the ingrained assumption that labor markets “should” clear. This creates an apparent need to explain the observation that they frequently do not do so. It leads to the situation where the aim of theory becomes to account for the divergence of reality from standard theory, rather than to explain reality itself.

(iii) Canonical search-and-matching theory

Another theory that attempts to explain non-frictional unemployment, the failure of the labor market to clear as predicted by standard theory, is search-and-matching theory. It suggests that equilibrium is delayed because the equilibration process is slower than traditional theory assumes. This is due to the difficulty that potential employees and workers have in finding
each other – a friction that means we never quite get to equilibrium. It implies that unemployment is just a question of reallocation – “the closeness of the match between the desired and actual characteristics of labor and capital inputs” – involving distance, skills, etc (Davis et al 1998, 106). In this perspective, “unemployment consists of workers who lose their jobs because it is not to their advantage (and to their employer’s advantage) to continue employed” (Pissarides, 2000, xvi), which is a poor account of, for example, plant closure.

Clearly, search-and-matching theory only applies once jobs have already been created, that is, after the investment decision and its announcement. It has nothing to say about firms’ decisions on scale, technology, location, closure, etc, nor about depressed regions or recessions, and is thus irrelevant to the various scenarios.

On the other hand, it is well suited to describing the dynamics of labor supply, and could thus be relevant to the tactical adjustments described above, that occur in the course of production. It therefore has a complementary role to the real investment led economy perspective proposed here.

(iv) John Maynard Keynes

The perspective closest to the real investment led economy viewpoint in the previous literature is that of Keynes in the General Theory:

“The unusual feature of Keynes’s analysis is its recognition that all the power is in the hands of producers. This is not because they occupy a monopsony position in the labour market, but arises simply from the temporal ordering of the process of producing for (uncertain, future) market sale: firms decide how much employment to offer on the basis of their expectations, and in the decentralised system of Western capitalism, if these decisions do not absorb all the labour available, that is just too bad” (Chick, 2007).

The parallels are clear: in both accounts, the firm is the initiator and this gives it a degree of power. The quantity of employment depends on the totality of investment. In addition, both theories involve time. In Keynes’ case, the wages are paid before the product is sold; with the real investment led economy viewpoint, the longer timescale of investment and its medium-term payoff is involved. Money is necessary to both perspectives. Uncertainty is central, because the product may not succeed in the market, which is accentuated in the real investment led economy view more than in that of Keynes, in view of the longer timescale and larger number of variables involved in investment.

The theories are fundamentally distinct, however, because Keynes was here referring to production, not investment. The time delay and the uncertainty refer to the gap between the payment of wages and the receipt of revenue, whereas with investment the intended payback period is spread over a number of years. And as is well known, Keynes’ theory of unemployment was quite different from that proposed here.

10. The real investment led economy explanation of each scenario

As previously stated, according to the real investment led economy view, firms invest when the expectation of future gains exceeds the expectation of the necessary costs by an amount
that generates a satisfactory return on capital. This clearly applies to setting up a new company, such as an airline. The types and quantities of jobs flow from the calculations that lie behind the business plan.

With technological change and relocation, the calculation is that a fundamental change is considered advisable, because they would generate a better investment. The costs of the innovatory investment, including managerial time and effort, are justifiable in view of the expectation of better performance in the medium and longer term. Also, innovative investment may be considered necessary in a dynamic sector, to avoid being displaced by competitors.

Understanding the other three scenarios requires a firm-centered perspective, with profitability at its center, if not necessarily investment as such. As stated above, plant closure is the mirror image of investment, which affects firms that are in a relatively weak position. The decision to discontinue production is taken by the firm, taking the economic conditions into account. This sets the agenda, and any other decisions then have a secondary status, including possible negotiations with workers or their representatives.

Similarly, in a depressed region or a recession, the firm’s profitability is crucial. With weak demand, costs may need to be reduced in the short term, which will often imply loss of jobs. The firm-centered view in this situation is again the mirror image of investment: a firm in a weak position, faced with difficult decisions, needs to focus on the future scale of operation, possibly reducing it so to minimize losses. It is true that a possible response would be to cut wages (and other costs), rather than to reduce the scale of production, and this has occasionally been known to occur. But usually it is the scale of production that is chosen, because the firm has a continuing relationship with its employees that would be impaired by wage cuts – its priority is the morale and therefore the productivity of those workers who remain employed. According to interviews with firms conducted after the early 1990s recession in the north-eastern United States, “resistance to pay reduction comes primarily from employers, not from workers or their representatives, though it is anticipation of negative employee reactions that make employers oppose wage cutting” (Bewley 1999). Note that this response was phrased in terms of firms’ decisions based on their perceptions of the situation, which accords with the perspective that firms take decisions in the light of their economic environment.

11. The real investment led economy view and puzzles in labor economics

Non-frictional unemployment

Perhaps the central puzzle of labor economics is the existence and persistence of unemployment, sometimes on a large scale. The real investment led economy viewpoint suggests that the explanatory focus should be on the opportunities for firms to make investments that promise to be profitable and thereby to create employment. There is no assumption that market forces “should” bring about a labor market equilibrium, particularly in the short term. The cause of non-frictional unemployment was outlined in section 5.

---

6 It is unclear how generalizable such findings are to other places and periods as the research has not been done.
Wage stickiness

The real investment led economy perspective proposes that the approximate wage level is set strategically at the time of investment, albeit that the precise wages may then be negotiated when employees are hired, and again when new matches are made, as well as being modified e.g. during annual pay negotiations. The rough level of wages then persists at least until the next investment is made, or until plant closure. Large changes are usually not made, as firms fear that this would destabilize the situation and might well reduce productivity and therefore profit. This reasoning is reinforced by the observation that contracting the volume of production saves not only the wage bill of the dismissed workers but also (in many cases) the other costs involved, such as raw materials and the costs of maintaining premises; in contrast, very large wage reductions would be needed to make significant savings (Bewley 1999) – as in the context of investment, the wage bill is only one item, albeit an important one, in the totality of the firm’s situation.

This would predict that real wages in capitalist firms are rather stable, with only rather minor adjustments being made between major investments, which is in accordance with the usual real-life situation. Wage stickiness is thus a consequence of the theory, not an assumption. Note also that this argument is distinct from the notion that fixed-length contracts reduce wage adjustment – in the real investment led economy view, the same approximate wage level is retained when one worker leaves and is replaced by another with similar skill level, although it may be modified by prevailing market conditions.

The wage flexibility puzzle and the unemployment volatility puzzle

The wage flexibility puzzle is the observation that the cyclical variation in wages is typically very low, whereas standard theory predicts that wage levels should fall in a recession (Pissarides 2009). Instead, employment usually fluctuates more than predicted – the unemployment volatility puzzle (Chodorow-Reich and Karabarbounis, 2013). The real investment led economy view predicts that wages are roughly stable between one investment and the next, because approximate wage levels are part of the firm’s investment decision. This reasoning is reinforced by the broader arguments given above, relating to the effects on morale of wage cuts compared with the scaling down of production, and to the relative savings from the two alternative courses of action.

Evidence on vacancies

According to the real investment led economy perspective, employment is created by firms – but not necessarily as abundantly as would be needed for everyone who is willing to work to be able to find a job. This means that there is typically (although not invariably) a shortage of jobs, of a magnitude that depends on how successful the economy is, in the sense of the firms being able to find potentially profitable investments. The clear prediction is that vacancies are scarce, and this accords well with the evidence: it has long been recognized that reported vacancy rates are low, that there are typically many applicants for each vacancy, and that vacancy durations are typically short. Classic studies include Holzer (1994) for the United States, Beaumont (1978) and Roper (1986; 1988) for the United Kingdom, and van Ours (1989) and van Ours and Ridder (1992) for the Netherlands. It also fits with the observation that job offers made by firms are usually accepted by a large majority of applicants, e.g. 90% (Barron et al., 1985; 1997).
In addition, this perspective fits well with the less-researched observations (what Manning (2003) calls “casual empiricism”) that it is difficult for unemployed workers to find employment, and that typically only a small proportion of firms report difficulties in hiring labor – although there may well be skill shortages due to a mismatch of desired and available skills. Nevertheless, the idea that there can be a shortage of jobs, e.g. in a recession, meets with resistance in some quarters. For example, Rogerson and Shimer (2011): “workers are constrained from working as much as they would like during recessions, perhaps because search frictions prevent them from finding a job”. This is despite the abundant evidence that recessions are characterized by a shortage of vacancies.

**The correlation of employers’ characteristics with wages**

The emphasis on the firm’s central position in making the main decisions on employment and wage levels naturally leads to the prediction that the characteristics of a particular firm will have important consequences for those decisions.

The best-known observation is the employer size-wage effect (ESWE), which has been documented across a large number of countries and across a variety of different types of sector (e.g. Adamczyk, 2015). A firm’s size “now” is largely a consequence of its degree of success in previous periods. It can therefore afford to pay higher wages, and that also means it can attract higher-quality workers, and is better able to retain them than the less successful firms (Brown and Medoff, 1989). Rather than firm size in itself being an explanation of the higher wage, both of these features result from the superior past performance of firms that are now large.7

In addition to ESWE, other characteristics of the firm have been shown to be correlated with wages (Manning, 2003). These include profits, profits per worker and productivity. In particular, inter-industry differences in wages have long been recognized empirically (Slichter, 1950; Krueger and Summers, 1987; Dickens and Katz, 1987; Krueger and Summers, 1988; Katz and Summers, 1989; Akerlof and Yellen, 1990). For example, in sectors with a relatively low proportion of costs that are due to wages, and corresponding higher costs of raw materials/fuel or capital, firms are able to pay higher rates – wages are less salient for them. And even the gender pay gap has an important firm-specific component (Card et al., 2014).

12. Conclusion

The account put forward in this paper is based on a description of the way that firms operate, notably the time order of decisions and events. In a sense the description is obvious, and probably most labor economists recognize these phenomena from everyday life. The issue is that the phenomena are not represented in the dominant theories – but they do make sense when integrated as the firm-centered perspective.

The real investment led economy view promises to be informative in the context of labor economics. Its account of firm decision making, emphasizing the strategic importance of the investment decision, is realistic. The focus on investment means that the theory is long-term, dynamic and forward looking. This perspective is able to explain non-frictional unemployment,

---

7 This does not mean that large firms are necessarily destined to out-perform their competitors in the future – size is a consequence of past performance, not an accurate predictor of future performance.
to provide a good account of a range of possible scenarios, and to shed light on some of the existing puzzles in labor economics.

We are now in a position to answer the questions posed at the beginning of the paper. Who creates jobs? Firms do. When? In the course of investment. Why? To engage in production, in order to make a profit. How? By using their buying power to attract and retain workers, to develop a relationship of authority with them, and thereby to direct their productive forces.

One of the implications is that individual workers’ job-search behavior primarily affects the allocation of particular workers to particular jobs, not the total number of jobs or their characteristics. An important point here is, however, that the economic environment of the firm – crucial in forming its investment decisions – includes the availability of workers with specific skills and other qualities, as well as the going wage rate for employing them. The role of potential workers in these decisions is thus seen as being at one remove, mediated through firm decision making, at the level of a group of prospective employees, and at the time of the firm’s investment decision when firms and potential workers may not yet have met. This implies a departure from a long tradition within labor economics, that seeks to explain the features of (un)employment by referring to workers’ preferences, e.g. their reference wage, and/or their job-search behavior; these latter factors however remain relevant to the tactical adjustment of wages and working conditions to economic conditions as they change, as well as to labor supply decisions, and to who gets which job.

The real investment led economy perspective explains the existence of persistent unemployment. It does not start from the assumption that the labor market should live up to its name – that the wage level “should” adjust, allowing an equilibrium to be reached. If one takes that conventional starting point, together with the observation that unemployment is common in the real world, one needs to explain the discrepancy. This runs the risk of providing an explanation that is totally implausible – e.g. the notion that something as small as “menu costs” can cause large-scale unemployment.

It also does not portray unemployment as either voluntary or involuntary. It is clearly not voluntary, in the sense that people choose not to work when the wage level falls below expectations. But also, it is not well characterized by the term “involuntary”, i.e. that a person is willing to work at the prevailing wage yet is unemployed, because that formulation suggests that to understand unemployment, one should focus on what unemployed people are prepared to do. Rather, it is a simple matter of the number of jobs created by firms, in comparison with the number of people who would like to work, stratified by skill and location.

The presentation of this perspective is not complete: for example, the relationship of the firm-centered view with other important topics, such as the role of government policies, remains to be analyzed. Its macroeconomic implications also remain to be developed, for example the impact of effective demand for firms’ products on their investment decisions. Other macro topics that have historically been considered important, but have not yet been considered in this context, are the possible existence of a natural rate of unemployment and whether this varies over time; and the relationship of the labor market with inflation. These are all left for future research.
References


Author contact: [m.joffe@imperial.ac.uk](mailto:m.joffe@imperial.ac.uk)

**SUGGESTED CITATION:**

You may post and read comments on this paper at [https://rwer.wordpress.com/comments-on-rwer-issue-no-83/](https://rwer.wordpress.com/comments-on-rwer-issue-no-83/).