WORKING PAPER 2020:13

Reciprocity and the interaction between the unemployed and the caseworker

Gerard J. van den Berg
Iris Kesternich
Gerrit Müller
Bettina M. Siflinger

IFAU
INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Reciprocity and the interaction between the unemployed and the caseworker[a]

Gerard J. van den Berg[b] **Iris Kesternich**[c]
Gerrit Müller[d] **Bettina M. Siflinger**[e]

May 25, 2020

Abstract

We investigate how negatively reciprocal traits of unemployed individuals interact with "sticks" policies imposing constraints on individual job search effort in the context of the German welfare system. For this we merge survey data of long-term unemployed individuals, containing indicators of reciprocity as a personality trait, to a unique set of register data on all unemployed coached by the same team of caseworkers and their treatments. We find that the combination of a higher negative reciprocity and a stricter regime have a negative interaction effect on search effort exerted by the unemployed. The results are stronger for males than for females. Stricter regimes may therefore drive long-term unemployed males with certain types of social preferences further away from the labor market.

Keywords: behavioral response, active labor market policy, monitoring, welfare, job search.
JEL codes: J64, D91, I38.

[a]*Acknowledgements*: We gratefully acknowledge financial support by the Thyssen Foundation (20.15.0.080WW), the Fonds Wetenschappelijk Onderzoek (FWO) through project grant G067718N, KU Leuven through C1-project grant C14/17/009, and the Deutsche Forschungsgemeinschaft (DFG) through project grant SI 2239/1-1. We thank the Research Department Panel Study Labour Market and Social Security of the IAB Nuremberg, in particular Mark Trappmann, for help with the survey data, and the Department of Data and IT-Management of the German Federal Employment Agency, in particular Markus Köhler, for providing the administrative data. We are grateful to Gudrun Fausel from ProIAB and Peter Kupka for expert qualitative information on interactions between unemployed and caseworkers. We thank Heiner Schumacher and Maria Hemström Hemmingsson for comments and suggestions.

[b]University of Bristol, University of Groningen, IFAU Uppsala, J-PAL, IZA, ZEW, CEPR.
E-mail: gerard.vandenberg@bristol.ac.uk
[c]KU Leuven. E-mail: iris.kesternich@kuleuven.be
[d]Institute for Employment Research. E-mail: gerrit.mueller@iab.de
[e]Tilburg University, Netspar, CESIfo. E-mail: b.m.siflinger@uvt.nl

1 Introduction

Active labor market policies (ALMP) are widely used to combat unemployment in developed countries. Large shares of the unemployed are enrolled in ALMP programs or exposed to such policies, and many OECD countries spend substantial parts of public budgets on them (Crépon and van den Berg, 2016). In general, unemployment benefits recipients are obliged to follow a rules on search effort, showing up at regular meetings at the employment agency, carrying out administrative duties and accepting job offers deemed suitable. The policies of monitoring these activities and imposing punitive sanctions to those who violate the rules is usually seen as one of the key ALMPs. These policies have been referred to as "sticks" policies (as opposed to "carrots" policies such as job search assistance; see e.g. Arni et al., 2015). While there is evidence that these increase re-employment probabilities (Crépon and van den Berg, 2016), they may be associated with unintended consequences such as lower re-employment wages or financial, emotional or health problems (van der Klaauw and van Ours, 2013; van den Berg and Vikström, 2014). Thus, sticks may drive individuals away from regular employment (van den Berg *et al.*, 2017).

In Germany, "sticks" have become an important part of unemployment policy after the so-called Hartz reforms. These were enacted between 2003 and 2005 and constitute one of the most ambitious labor market reforms in Germany after World War II. The aim was to make the German labor market more flexible and to reduce unemployment durations (Fahr and Sunde, 2009). The so-called Hartz IV package restructured the unemployment benefit system in such a way that the interaction between the caseworker and the unemployed is now determined by explicit monetary and non-monetary incentives. While uncommon in the old system, sanctions for low engagement in search activities are frequently applied under Hartz IV (Jacobi and Kluve, 2007). Most OECD countries employ a similar policy mix of supportive and restrictive policies (OECD, 2013).

Previous studies document heterogeneity in how stringent caseworkers are in their monitoring and in handing out sanctions (van den Berg *et al.*, 2014b; Arni *et al.*, 2015). This reflects that caseworkers have discretionary powers in their application of the rules. The heterogeneity can be captured as caseworker-specific regimes (Arni *et al.*, 2015; Behncke *et al.*, 2010b). In this study, we document heterogeneity in reactions of unemployed individuals to the strictness of "sticks" regimes applied by caseworkers, and we show that this heterogeneity is driven by negative reciprocity. Reciprocity is a social preference and can be defined as the degree to which an individual is willing to incur costs in order to reward positive or punish negative behavior by others (Falk and Fischbacher, 2006). It plays an important role in labor markets, where negative reciprocity tends to reduce effort (Dohmen *et al.*, 2009). Fehr and

Gächter (1998) and Fehr *et al.* (1997) show that introducing explicit incentives can even be harmful with reciprocal workers. Thus, taking into account reciprocal motives is important for predicting and understanding the response to policies, in particular in labor markets, where contracts cannot fully specify all important aspects of a relation between two parties (Gächter and Fehr, 2008). Consequently, sticks might not have the desired effect if unemployed individuals retaliate in response to a perceived wrong and decline to cooperate with the labor market agency even if non-cooperation is costly for themselves (thus showing negative reciprocity). In our paper we focus on unemployed welfare recipients. These are often long-term unemployed and are at risk of an increasing detachment from mainstream society. It is an important policy question whether "sticks" policy regimes perhaps unintentionally serve to drive at least a fraction of them further into that direction.

For our empirical analysis we collect measures of reciprocity as part of the German "Panel Study Labour Market and Social Security" (PASS). This is an annual longitudinal panel survey which aims at analyzing the dynamics of unemployment benefits after the introduction of Hartz IV (Trappmann *et al.*, 2019). PASS contains a random sample of long-term unemployed welfare recipients (officially: Unemployment Benefits Type II), drawn from the unemployment register of Federal Employment Agency statistics. It collects detailed information on respondents' labor market histories, including measures on search efforts and search motivation, which we will use as our outcome measures.

To obtain a measure of the strictness of the regime applied by a caseworker, we use unique administrative records stored in the internal IT system of the Federal Employment Agency. These contain information on treatments received by unemployed individuals as well as identifiers of the team to which the caseworker belongs. Specifically, we exploit data on the share of "Integration Agreements" (IA) that are enforced by caseworkers through an administrative legal act. As we explain in more detail below, an IA is a written contract that stipulates rights and obligations of unemployment benefits recipients. Its contents is based on a template that may slightly vary across occupation and family status but that is otherwise uniformly specified. This template reflects existing rules and laws (see e.g. Schütz et al., 2011, and Boockmann et al., 2013, for descriptions of the IA). In fact, the contents of the IA is mostly a list of obligatory job search activities on the part of the unemployed worker, such as a minimum number of job applications per time unit. The first signing of this contract takes place upon entry into the relevant benefits scheme, at the end of the first meeting of the unemployed and his/her caseworker. The IA is a central element of the incentive setting under Hartz IV and is the basis of the enactment of the monitoring and sanctions policy.[1]

[1]Policy instruments that are similar to (but weaker than) the IA are used in other countries, for example Austria, Denmark, and the Netherlands (van den Berg *et al.*, 2014a).

Both the benefits recipient and the caseworker should sign the IA. However, the caseworker may impose the IA unilaterally if the unemployed refuses to sign. In the latter case the signature or consent of the unemployed is not warranted. The latter way of imposing the IA is called an "administrative act". In either case, the contract is legally binding.

A single caseworker can coach up to about 60 unemployed clients at any given point in time. Each caseworker is a member of a caseworker team. A team can contain up to 20 full-time or part-time employed caseworkers and each local employment agency can house a number of such teams. The team is the natural organizational unit to coordinate work and the implementation of policy regimes. A regime in which the team of caseworkers applies relatively strict "sticks" policies will be a regime in which it relatively often enforces IAs through administrative acts. We do not directly observe the former but we can use the degree of enforcement through administrative acts by the team as a proxy for the strictness of the regime.[2] This is a contextual indicator of the strictness conditions facing the unemployed individual.

To implement this empirically, a number of issues need to be addressed. First, to rule out reverse causality, the individual's team-level measure for strictness of the regime excludes whether the individual him- or herself was subjected to an administrative act. Second, we need to address the composition of clients within caseworker teams and the assignment of clients to teams. Here we exploit the quasi-random assignment of unemployed to caseworkers. As a general rule, unemployed individuals are matched to caseworker teams in alphabetical order of the name of the former, with different teams covering different sets of letters of the alphabet, or on a first-come first-serve base.[3] This supports the notion that individuals are not systematically assigned to specific teams of caseworkers based on specific personality traits or labor market characteristics of the former. To further reduce selectivity biases, we condition on an unusually rich set of characteristics of the employment agency ("job center"), of the team within the center that the unemployed is assigned to, and of individual-level characteristics. This includes information on the outcomes of profiling of the unemployed, summarizing their perceived chances in the labor market. At the level of the job centers themselves, we have unique access to qualifiers that are used in the internal quality benchmarking by the Federal Employment Agency. The latter information as well as the identifiers of caseworker teams have not been used before in any academic paper.

[2]Using indicators based on sanctions has the disadvantage that sanctions are given strictly after entry into the benefits scheme and are only observed if they occur before re-employment. This creates a dependent competing risks setting with identification challenges.

[3]This information is provided to us by experts at the IAB who have carried out qualitative interviews among employees at employment agencies. Our data do not contain the names of the unemployed individuals or the time of day at which meetings were held, so that we cannot verify this independently.

As a third issue, we need to address implications of the fact that if an unemployed individual has an informed view of the strictness of the caseworker team regime before deciding whether to sign the IA or not then this may influence the outcome of that decision. In general there is a cost involved in being the subject of an administrative act, namely that one may subsequently be labeled as non-cooperative and be exposed to more frequent and/or harsher monitoring, compared to when the IA remains unenforced or is open to further discussion. This has a specific implication for the association between the strictness of the regime and the rate at which administrative acts are handed out. If the regime is very relaxed then only few administrative acts are given per time unit. If the regime is somewhat less relaxed then, most likely, more of these acts will be given. But if the regime is perceived to be extremely strict then the unemployed will tend to sign any IA, so that ultimately only few administrative acts are given per time unit. Thus, the association between strictness of the regime and the rate at which administrative acts are handed out can be non-monotonic. This in turn implies that if the range of regimes spans the full spectrum then the fraction of administratrive acts may not be informative on the strictness of the regime. Such extremely strict regimes may not be very relevant in practice,[4] and the issue would arguably be more problematic if the frequency of imposed sanctions is used instead of the the frequency of administrative acts, since the cost of receiving a sanction is higher. Still, we need to rule out settings where at some teams the deterrent effects of administrative acts are known to be so uniformly high (in terms of detection, enforcement and subsequent monitoring and punishments) that these acts effectively never need to be applied to the clients of that team.

Our main result is that negative reciprocity of unemployed clients creates heterogeneity in how the unemployed react to stricter regimes. Specifically, more negatively reciprocal individuals significantly decrease the effort invested in job search in response to stricter regimes. For instance, if the share of integration agreements enforced increases by one standard deviation, individuals who are about one standard deviation more negatively reciprocal than the average reduce their probability of searching for a job by 2.5 percentage points.

The results suggest that a "sticks" policy can drive long-term unemployed individuals further away from the labor market, where this effect is driven by individual heterogeneity in social preferences. Interestingly, the strong interaction effect between negative reciprocity and stricter regimes is mainly driven by men. Note that the long-term unemployed constitute a vulnerable population. The individuals at risk may become so detached from the regular labor market in response to "sticks" that they will not be able anymore to escape the trap of social isolation and/or the loss of attachment to mainstream society. While the impacts of financial

[4]Again, this is confirmed by experts at the IAB who have carried out qualitative interviews among caseworkers and other employees at job centers.

incentives on unemployed individuals are relatively well-understood, our study reveals that psychological factors and preferences that may interact with incentives are important as well. Policy makers may consider ALMP tools that allow a more flexible individual application. To this end, one might even consider to tailor labor market policies more to the unemployed individuals' personalities.

Our paper is connected to the literature on ex ante effects of possible future treatments on not-yet-treated individuals, such as sanctions in a monitoring scheme (see e.g. Black et al., 2003, and Arni et al., 2015). In this literature, the standard explanatory line of reasoning is that restrictive sticks policy regimes induce individuals to search for work more actively and reduce their reservation wage, thus increasing re-employment rates. Yet other studies have examined effects of warnings or notifications of the likelihood of future individual treatments (see Lalive et al., 2005, for a "sticks" policy).

We also contribute to recent literature on the interaction between the caseworker and the unemployed. Behncke *et al.* (2010a) and Behncke *et al.* (2010b) show that caseworkers who belong to the same social group as the unemployed individual increase re-employment probabilities, and that having a less cooperative caseworker increases re-employment probabilities. Our paper complements these studies by explicitly considering heterogeneity in how unemployed individuals react to the strictness of the regime.

Finally, we add to the literature on social preferences and behavioral labor economics by showing that negative reciprocity plays an important role in how the unemployed react to negative incentives. While the correlation between negative reciprocity and unemployment incidence has been established in Dohmen *et al.* (2009), we are the first providing causal evidence.

The outline of the paper is as follows. Section 2 discusses briefly the behavioral concept of reciprocity in labor economics. Section 3 provides some background information on the institutional context and discusses the data. The empirical strategy is presented in Section 4 and the empirical results are in Section 5. Section 6 concludes.

2 Reciprocity

In the context of the labor markets the terms "reciprocity" and "gift exchange" are often used interchangeably. The seminal paper on "norm-gift exchange" in the labor market is Akerlof (1982). In his model workers respond to wages that are larger than the market-clearing wage by providing more effort. They do this, because their behavior is governed by norms of gift exchange – thus they reciprocate to the "gift" of high wages by providing high effort. Reciprocity is used to explain why firms pay higher wages than the market-clearing

or minimum wage (to induce efforts); why there can be involuntary unemployment (because firms choose to pay wages that are higher than the market-clearing wage); why there is no wage cutting in recessions (to avoid negative responses from reciprocal workers); and why explicit performance-based pay is so rare in actual labor markets (because reciprocity and monetary incentives are substitutes).[5]

Reciprocity, or a general cooperative attitude, mainly plays a role when contracts are incomplete and when there is considerable discretion over effort (Fehr and Gächter, 1998). Fehr and Gächter (1998) and Fehr *et al.* (1997) indicate that introducing explicit incentives can even be harmful with reciprocal workers. Hartz IV has introduced both positive and negative economic incentives, and they even make these incentives more explicit by requiring a written contract between the caseworker and the unemployed. The literature on reciprocity suggests that, at least for strongly reciprocal agents, this might have unintended (negative) consequences. In our setting, negative reciprocity may play out as follows. If the unemployed individual stops providing search effort, this will not only be costly and harmful to himself, but also to the caseworkers who have to meet certain targets with respect to the successful reintegration of their unemployed clients. By ceasing to cooperate, the unemployed breaks the implicit contract between unemployed individuals and society that is central to Hartz IV – namely that the state supports individuals with unemployment benefits in exchange for search effort.

The main evidence on the impact of reciprocity on the labor market comes from incentivized experiments (see the literature summaries by Fehr and Gächter, 1998; Fehr and Schmidt, 2006). Positive reciprocity has been established through so-called trust or gift exchange games (Fehr *et al.*, 1993) or in experimental labor markets in which workers reciprocate high wages with high effort (Gintis *et al.*, 2003). In a trust or gift exchange game, a sender can decide to keep his endowment to himself or send it to a receiver. The amount sent has some return, thus what the receiver gets is a multiple of what the sender sent. The receiver positive reciprocity is elicited by sending back a part of what he received from the sender. Negative reciprocity can be measured by using ultimatum bargaining games or in other types of games when respondents reject offers that they perceive as unfair or behavior that they consider as unfair at personal cost (Güth *et al.*, 1982; Carpenter and Seki, 2005). These experimental games have been played over and over again with very similar results in different places, at different times, with low and high stakes, in one-shot and repeated games, and using different subject pools (see Gächter and Fehr, 2008; Fehr and Schmidt, 2006; Camerer and Thaler, 1995). Reciprocity has also shown to be important for sustaining long-term relationships (Brown *et al.*, 2004). Thus, the importance of both positive and

[5]See Akerlof (1982, 1984); Fehr *et al.* (1993); Englmaier and Leider (2012).

negative reciprocity for equilibrium of stylized labor markets or cooperative games in experimental settings has been established convincingly. These studies also show that there is great heterogeneity with respect to reciprocity. In experimental labor markets only 40 to 50 percent of subjects exhibit a reciprocal effort pattern, but this is large enough a fraction to make it profitable for employers to pay wages above the lowest feasible wage (Fehr and Schmidt, 2006).

To what extent do these findings from the laboratory carry over to real labor markets? Gneezy and List (2006) conduct a field experiment in which workers are hired to code books in a library or to raise funds. They find that workers who are offered a "gift" in the form of a higher than the agreed wage show an increase in productivity. Gilchrist *et al.* (2016) differentiate between high wages and an unexpected unconditional pay rise and find that the latter again leads to increases in productivity. In Carpenter and Seki (2005) find that Japanese fishers who are more reciprocal also provide more effort in their actual work. Bewley (1995, 1994) conducted interviews with managers who asserted that the fear of negative reactions by workers is the main reason not to cut wages. Firing workers is mainly used to get rid of non-cooperative workers in order to sustain a cooperative attitude. Bellemare and Shearer (2009) investigate worker's response to a monetary gift by their employer, a tree-planting firm. They find that workers indeed increase their effort levels in response to the gift. Kube *et al.* (2013) document persistent negative effects of a wage cut on worker productivity in a field experiment. There is little evidence based on large random samples of the population. Dohmen *et al.* (2009) introduce a battery of survey questions on reciprocity into the German Socio-Economic Panel. On a methodological level, they show that positive and negative reciprocity are two different traits, which are only weakly correlated. Further, they show that more positively reciprocal individuals are more likely to be employed, are paid higher wages and provide more effort. In the paper most closely related to our research, Montizaan *et al.* (2016) link survey and administrative data to show that more negatively reciprocal workers reduce their work motivation more in response to a pension reform that curtails their pension rights. While Dohmen *et al.* (2009) show that negative reciprocity is correlated with being unemployed, to the best of our knowledge the channels linking (negative) reciprocity and unemployment have not yet been investigated. With this paper, we contribute to closing this gap in the literature by investigating how negative reciprocity changes the way in which the unemployed react to negative incentives. If negatively reciprocal individuals who are long-term unemployed decrease the effort invested in job search in response to "sticks" policies, this means that not only are negatively reciprocal individuals more likely to become unemployed, they are also more likely to be driven even further away from the labor market once unemployed.

3 Institutions and data

3.1 Institutional Background of the Hartz Reforms

In the spring of 2002, the German government formed a commission of politicians and business professionals to work out suggestions for a policy reform to overcome the high unemployment rates and to reduce the exorbitant costs of the social security system (Fahr and Sunde, 2009). The result was a series of comprehensive reforms of the social security system, the so-called Hartz reforms, consisting of four distinct laws (Hartz I–Hartz IV), which were implemented between 2003 and 2005. The Hartz IV reform created a new type of benefits, "type II benefits" (or SGB II) for which individuals can be eligible if they are unemployed but not eligible (or not eligible anymore) to regular unemployment insurance benefits (UI, or "type I benefits"). Receipt of type II benefits entails exposure to activation through explicit incentives (see Krebs and Scheffell (2013)). As a rule, individuals who enter unemployment from employment receive type I benefits for the first 6-12 months, depending on their prior employment history, and may receive type II benefits afterwards. (For unemployed above 50 the maximum possible UI entitlement period is longer, up to 24 months for those aged above 58.) While type I benefits depend on previous wages, type II benefits are flat-rate and means-tested and depend on the household composition. In our study, we focus on recipients of type II benefits. Note that this means that we do not consider short-term unemployed with a high likelihood of swift re-employment. For ease of exposition we refer to type II benefits as welfare and to its recipients as long-term unemployed.[6]

Officially unemployed individuals are called "clients" or "customers" who are entitled to professional, reliable and friendly service. However, the relationship between the caseworker and the unemployed is different from that between a service provider and a paying client. Since caseworkers have such far-reaching competences and margins of discretion, the relationship between the two parties is of a hierarchical nature (Baethge-Kinsky *et al.*, 2007). Unemployed individuals are subject to activation policies according to the principle of "assist and demand" (Jacobi and Kluve, 2007). On the "demand" side, the unemployed have to prove that they spend significant effort on finding a job and they have to be flexible in accepting job offers. In case the unemployed do not meet their obligations, caseworkers can sanction by reducing benefits. Caseworkers conceive of their role as activating the unemployed (Baethge-Kinsky *et al.*, 2007), and sanctions for low engagement in job search

[6]Long-term unemployment is often used to denote individuals with elapsed durations exceeding a year, whereas in our case some recipients of type II benefits have a shorter elapsed duration and some older individuals with durations between 12 and 24 months do not yet qualify for type II benefits. In this sense we use "long-term unemployment" as synonymous with being unemployed and receiving welfare.

activities, which were uncommon before the Hartz reform, are now frequently used (Jacobi and Kluve, 2007). Benefit receipt is now conditional on individual search activities and cooperation in the job finding process (Eichhorst *et al.*, 2010).

An individual who wants to claim welfare and who is able to work for at least 15 hours per week has to register at the local job center and is then assigned to a team of caseworkers. The first contact is made through the administrative staff at the reception desk on a first-come-first-serve basis (Kaltenborn *et al.*, 2011). Next, the unemployed individual is assigned to a caseworker team. As explained in Section 1, as a matter of principle the unemployed individuals are assigned to teams in alphabetic order of the name of the former, with different teams covering different sets of letters of the alphabet, or, alternatively, assignment is on a first-come first-serve base. Each team consists of about 10 – 20 caseworkers.[7] In principle, each unemployed person is assigned to a single caseworker within the team, again on a first-come first-serve base. Changes in caseworker (to another caseworker from the same team) occur only in special cases such as sickness or job changes of the caseworker. Teams meet regularly to discuss cases and to coordinate discretionary decisions and the implementation of policy regimes.

During their first meeting, the unemployed and the caseworker work out the strengths and weaknesses of the unemployed individual and develop specific goals towards reintegration into the labor market. This process is highly standardized and serves for the caseworker as a tool to assign the unemployed to one of six *profiles* with differing re-integration chances. The profile of the unemployed affects access to labor market policy measures such as training that an individual receives while being unemployed (Jacobi and Kluve, 2007). In descending order of their re-integration probability, it distinguishes between market, activation, assist, development, stabilization, and support profiles. For ease of exposition we aggregate the development, stabilization and support profile categories into a single category called complex profile. An unemployed person with a market profile does not require major support and has a perceived re-integration probability of close to one within six months of unemployment. By contrast, an individual with a complex profile has a very low perceived probability of employment within 12 months.

Towards the end of the meeting, the obligations of both, the caseworker as well as the unemployed, are summed up in the IA ("Eingliederungsvereinbarung") which is a written contract between the employment agency, represented by the caseworker, and the unemployed. Recall from Section 1 above that an IA that is mostly a list of obligatory job search activities on the part of the unemployed worker. Part of it outlines the legal impacts of the

[7]At each job center, there are separate teams for unemployed individuals under 25 and for those with higher education such as academics (Bender *et al.*, 2006).

agreement, for example, that sanctions apply in case the unemployed individual does not follow what is agreed upon (Kaltenborn *et al.*, 2011; Jacobi and Kluve, 2007). Both the benefits recipient and the caseworker should sign the IA. However, the caseworker may impose the IA unilaterally if the unemployed refuses to sign. In the latter case the signature or consent of the unemployed is not warranted. In either case, the contract is legally binding. Thus, while in theory the IA is two-sided, in reality it is an asymmetric contract (Bartelheimer, 2008).

The contract plays a central role in the relationship between the unemployed and the caseworker. The contract has to be renewed every six months, and is in practice renewed at about two thirds of meetings between the unemployed and the caseworker (Kaltenborn *et al.*, 2011). While the existence of an integration agreement is required by law, it is not required that the unemployed individual signs the contract for it to become valid. In fact, the caseworker can decide to enforce the integration agreement through an administrative act ("Verwaltungsakt") against the will of the unemployed. Caseworkers can also choose to use an administrative act when they want to make salient that the integration agreement is a legally binding contract, for example, when the unemployed does not show sufficient effort. This contributes to the asymmetric nature of the contract between the unemployed and the caseworker.

Employment agencies are subject to some organizational controlling and self-monitoring. To this end, there is a system of internal benchmarking. In order to compare the performance of different job centers, differences in economic and social structures between the regions that the job centers operate in are taken into account (Dauth *et al.*, 2013). The Federal Employment Agency categorizes each job center into one of 15 possible types which we will take into account in the empirical analysis.[8]

3.2 The survey data

We use individual level data from the "Panel Study Labour Market and Social Security" (PASS, DOI: 10.5164/ IAB.FDZD.1806.en.v1). The present subsection describes these data.

[8]For this classification, a large number of characteristics are used which come from the following broader categories: Unemployment and underemployment (e.g. un-/underemployment rates for different groups, seasonal unemployment); capacity of local labor market for job seekers receiving unemployment II benefits (e.g. share low/high qualified; degree of industrialization; employment trend); structural features of the group of individuals entitled to benefits and fit for work (e.g. total number, share above age of 55, share high/low educated unemployed, share migrants); structural features of benefit units (e.g. share single parents, units with more than 5 people); population and social structure (e.g. population density, population growth); and geographical characteristics of the job center (e.g. urban/rural, commuter friendly, infrastructure). For more detailed information on all variables entering the classification as well as the classification procedure see Dauth *et al.* (2013).

In the next subsection we discuss the merging of these survey data with administrative register data on caseworker teams.

PASS was initiated in 2006 to analyze the dynamics and long-term consequences of welfare benefits after the introduction of Hartz IV in 2005 (Trappmann *et al.* (2019)). Subsequently, PASS has been collected on an annual basis. The sampling of households comprises two sub-samples. One is representative for the German population. The other one is a random draw from the benefit recipient register at the Federal Employment Agency. Individuals are followed over time, even if they have left the original household, and partial refreshment of the sample occurs annually. Person-level interviews are conducted with all members of the sample household aged 15 or older. The data set is designed to provide a detailed picture of the process of reintegration of the unemployed and the quality of contact between employment agencies and unemployed. PASS contains information about respondents' socio-economic status, age, education and family structure. An advantage over other surveys is that it has a special focus on job search, measuring its different dimensions.

For the empirical analysis we use the stock of welfare recipients not currently in school, in wave 9 of the PASS (collected between February and September 2015). Among the latter we only consider individuals who are obliged to search for a job (and thus not exempted for reasons of illness, caregiving etc.) and who agreed to a link of the survey information to administrative data. As a result, our sample consists of 822 individuals. Somewhat loosely, this can be seen as constituting a stock sample of long-term unemployed individuals.

As outcome variables we use the following five measures of an individual's job search behavior,

- *job search: yes* is a binary measure indicating whether an individual has searched for a job during the last four weeks;
- *log(hrs spent on job search)* refers to the log hours an individual spent on average on job search in the past four weeks (where in fact we use (hours + one) instead of hours in order to deal with zero hours);
- *search motivation* is a binary measure that takes the value one of the individual's search motivation is very high, high or medium (corresponding to values 3-5 of the original variable) and zero otherwise (values 1-2);
- *number search activities* refers to the number of job search activities of an individual (summed over the number of times an individual has responded to job ads, posted ads, directly asked companies, and spontaneously submitted applications);

- *willingness to concede: yes* measures whether an individual agreed to concede on at least one of seven items (commuting one hour or more; low income; unfavorable working hours; job below skill level; strains at job; moving; long distance relationship).

For our study, we added survey questions eliciting reciprocity to the 9^{th} wave of PASS. The reciprocity measure is based on a set of questions developed by Perugini *et al.* (2003) and implemented in the German Socio-Economic Panel in 2005 by Dohmen *et al.* (2009). This measure has been experimentally validated using ultimatum games (Perugini *et al.*, 2003) and has been shown to correlate with behavior in the field in a way consistent with theoretical predictions (Dohmen *et al.* (2009), and Montizaan *et al.* (2016)). The measure of negative reciprocity takes the mean over how strongly a respondent identifies with three statements. These are: (1) "If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the costs"; (2) "If somebody puts me in a difficult position, I will do the same to him/her"; (3) "If somebody offends me, I will offend him/her in return". Responses are measured on a four-point Likert Scale where 1 means "does not apply to me at all", and 4 means "perfectly applies to me".[9]

Figure 1 shows the distribution of the measure for negative reciprocity.[10] The distribution is similar to that reported in Dohmen *et al.* (2009) for respondents of the GSOEP. There is a spike at 1 (not at all negatively reciprocal), and the distribution is skewed to the right with many individuals being somewhat negatively reciprocal and few individuals being very negatively reciprocal. The mean of the distribution is at about 2, and thus individuals stating that the statements describing negative reciprocity "do rather not apply to them". About 35 percent respondents in our sample have values larger than 2 and thus agree on average with the statement on negative reciprocity.

The sample at our disposal is essentially cross-sectional and hence does not allow for analyses of effects on unemployment durations. Inclusion of subsequent PASS waves (to the extent that these are currently available) would provide observations of duration outcomes. However, low exit rates out of welfare imply that only few durations are observed to be completed. Hence, an analysis based on our sample of 822 individuals would suffer from lack of power. Moreover, caseworker regimes may not only affect unemployment durations through search effort but also in different ways, complicating the interpretation of results. We view this as a topic for future research.[11]

[9] We recode the original survey question that was measured on the reverse scale.

[10] A few individuals have a mean value of negative reciprocity of 1.5, 2.5 and 3.5. This is due to the fact that few (18) individuals only answered 2 instead of 3 reciprocity items.

[11] A duration analysis can also be used to shed light on dynamic selection of unobserved confounders affecting distributions of covariates in stock samples; see e.g. Ridder (1984) and van den Berg (2001).

3.3 The administrative register data

The administrative data we use come from the internal IT system of the Federal Employment Agency for employment services and vocational counseling (VerBIS). VerBIS is the central tool used by the Federal Employment Agency in order to document processes, to exchange information, to check the correctness of the information, to steer processes, and to ensure compliance with the German law (Kaltenborn *et al.*, 2011). VerBIS contains all relevant information about the unemployed individuals (for example the outcome of the profiling process) and documents the counseling process (for example the existence of integration agreements). The data base we use is the data warehouse of the Federal Employment Agency as of March 2017 (DHW-Schicht AST, FST and LST SGBII, 2016). For each PASS respondent who agreed to the link with administrative data, we searched for the team of the caseworker they were counseled by. For each identified team, we extracted team-level information such as number of customers coached, office days, share of integration agreements enforced, share of male customers, share of unemployed individuals that were profiled as difficult to re-integrate etc. Not surprisingly, since we consider long-term unemployed welfare recipients, the complex profile is the dominant profile in the data. A separate category are unemployed individuals who could not be assigned to one of the profiles yet. We further create 15 dummies for the job-center types in the internal benchmarking.

As explained in Section 1, we define a specific measure for the strictness of the regime applied by a caseworker, based on "administrative acts" for non-signing an IA. This excludes the IA of the unemployed individual him/herself. Across the 822 unemployed individuals in the merged data, on average about 3.4% of all the other IAs handled by the individual's caseworker team are enforced through administrative acts. In fact, it makes more sense to define the share of enforced IAs as a fraction of the number of office days of the team. This increases the average share of enforced IAs to about 7%. In a robustness check, we further redefine the share as the share per 100 clients per team. Descriptive statistics for all variables are in Table 1.

4 Empirical strategy

To address our research question on the importance of reciprocity in the interaction between the caseworker and the unemployed as well as its consequences on search outcomes, we estimate the following baseline equation for each of the outcome measures listed in Subsection 3.2,

$$Y = \beta_0 + \beta_1 NEGREC + \beta_2 AGRENF + \beta_3 (AGRENF \times NEGREC) + X\Gamma + \varepsilon \qquad (1)$$

where Y denotes the relevant measure of search effort, $NEGREC$ is negative reciprocity, $AGRENF$ is the indicator of the strictness of the caseworker's policy regime, and ($AGRENF \times NEGREC$) is the interaction term. The vector X comprises of the control variables while ε is the error term. We are interested in how the reaction of the unemployed individual's search efforts to the strictness of the regime depends on negative reciprocity. The coefficient of interest is therefore β_3, as this coefficient allows us to capture heterogeneity in the reaction to stricter regimes by reciprocal preferences. For ease of interpretation we standardize the variables $AGRENF$ and $NEGREC$. Thus, the coefficients β_1, β_2 and β_3 are interpreted as changes in outcomes generated by a change in mean $NEGREC$ and/or mean $AGRENF$ by one standard deviation. We estimate equation (1) by OLS, clustering standard errors at the team level.

We follow the literature on reciprocity by regarding reciprocity as a social preference that develops during childhood and youth and remains relatively stable during adulthood and until old age (Gutiérrez-Roig *et al.*, 2014). Thus, we take it that reciprocal preferences are not affected by the regime applied by the caseworker. As measure of the strictness of the caseworker's regime we use the average rate of administrative acts at the caseworker-team level per office day, excluding whether the client him- or herself was subjected to an administrative act. As explained in Section 1, this measure has a number of advantages[12] and is immune to a number of endogeneity concerns. Notably, its usage is supported by the quasi-random assignment of clients to teams and the controlling for an unusually rich set of covariates.[13] Specifically, regarding X, we distinguish between covariates measured at the individual level, at the team level, and at the job-center level (see also Table 1). The former include individual background characteristics such as gender, age, migration status, education, number of kids in the household, physical and mental health, disability and household composition. Covariates measured at the team level include features of the distribution of clients per team, such as the share of males and the fractions of younger and older unemployed and the fractions with a migration background and being a single parent.

[12]This includes aspects of the timing of the IA. Recall from Section 1 that usage requires us to rule out admittedly unrealistic settings where at some teams the deterrent effects of administrative acts are known to be so uniformly high (in terms of detection, enforcement and subsequent monitoring and punishments) that these acts effectively never need to be applied to the clients of that team.

[13]For completeness we also mention that the individual's own receipt of a sticks treatment is endogenous, since receiving such a treatment is likely to be associated with unobserved characteristics that determine individual search effort.

We also include the share of unemployed having a complex profile, are not classified yet or do not need to be profiled. Covariates at the job-center level include the classifications that are used for internal benchmarking of the centers. These classifications measure in which environment the job centers operate, combining factors like the regional economic and social conditions. They therefore also describe local labor market conditions. Jointly, these covariates capture economic conditions and the type of unemployed coached by the team, presumably affecting both search effort and the regime.

5 Results

5.1 Team-level assignment

As discussed above, our identification strategy relies on the quasi-random assignment of unemployed individuals to teams of caseworkers. Since the matching of unemployed to teams is generally based on alphabetical order, the assignment should be random. Accordingly, all teams should have a similar composition of unemployed individuals, and they thus should show a similar share of integration agreements enforced through an administrative act. To support this conjecture, we regress the (non-standardized) mean receipt of enforced integration agreements on our team-level characteristics. Column (1) in Table 4 displays the corresponding regression results without additionally controlling for job center characteristics. In total, we have 477 teams to which the 822 individuals were assigned to. On average, each team in our sample supervise a bit less than two unemployed individuals. In Column (1), none of the estimated coefficients is statistically significant, indicating no systematic variation in the share of integration agreements enforced along specific team-level characteristics. In Column (2) of Table 4 we additionally control for job center specific characteristics. Column (2) of Table 4 shows the estimated coefficients when fixed effects for the job center types are included in the regression. While many estimated coefficients increase in magnitude, the negative correlation between the share of clients being above the age of 55 and the share of integration agreements enforced through an administrative act at the team level becomes statistically significant on the 1 percent level. This coefficient may be an artifact of the chosen reference category for the job center fixed effects. The reference is "type Ia" which refers, among others, to job centers with a high fraction of unemployed individuals aged 50+. The estimated coefficient thus seems to capture the correlation between the job center type rather than the team-level characteristics. In any case, we control for both the share of individuals aged 50+ and the job center type in our main specification. Overall, the results from the team-level regression show little variation in the share of the integration

agreement enforced by administrative act by observable characteristics. This provides some support for our assumption that the matching of teams and individuals is quasi-random.

Quasi-random assignment is further supported if the correlation between individual negative reciprocity and our measure on the strictness of the regime is close to zero. Zero correlation would speak against issues of selection in which more negatively reciprocal individuals are supervised by teams using stricter regimes. It also may countervail the opposite hypothesis, namely that the strictness of the regime influences the individual level of negative reciprocity. Table 3 displays the results of regressing negative reciprocity on the share of integration agreements enforced through an administrative act. Across all specifications, the correlation is close to zero and not statistically significant. While we are aware that more periods would be required to neglect any (dynamic) selection, this finding supports our identification strategy.

One further issue is whether the type of regime applied by a team of caseworkers is stable over time or whether it reacts strongly to changes in labor market conditions. The team-level information on the share of integration agreements enforced by an administrative act is also available for the years 2012–2014. Thus, we can assess whether our "sticks" treatment has been commonly used before 2015 and whether the share of teams using sticks treatments changed over time. Table 2 displays different measures of the time trend in the share of integration agreements enforced for 2012–2015. The correlation coefficients across years in Panel A are all above 0.8, indicating a strong persistence in using our "sticks" treatment. This persistence is corroborated by the summary statistics presented in Panel B. The average share of integration agreements enforced by an administrative act was about 3.2 percent in 2015, and this fraction remains stable when going back in time. Table 2 provides us with a good indication on the stability of using administrative acts as an instrument to enforce integration agreements.

5.2 Estimates for the main specification

In this subsection we discuss the results for our main specification. Table 5 shows the estimated coefficients when measures of search effort are regressed on the severity of the regime, negative reciprocity, and their interaction. All our outcome variables are coded so that a higher value is associated with stronger search motivation, a higher search intensity, or a stronger willingness to concede. We standardized our main independent variables, the measures for negative reciprocity and for the share of integration agreements enforced through an administrative act among all contact days of the team. This allows us to interpret coefficients in terms of changes in standard deviations from mean values.

The general pattern is as follows. The main effect of facing a stricter regime is positive, but not statistically significant. Being more negative reciprocal reduces search effort for all outcomes. These effects are statistically significant for job search, log hours searched and the willingness to concede. Specifically, an increase in the strictness of the regime by one standard deviation, that is by about 10.6 percent, has a positive, but not statistically significant effect on all measures of search effort. Starting from a mean value of about 2, an increase in negative reciprocity by one standard deviation (about 0.8) significantly reduces an individual's search effort for all measures but the search motivation and the number of search activities. For job search (Column (1) in Table 5), this implies that for a mean share of enforced integration agreements of about 6.6 percent among all contact days per team, individuals who are somewhat negative reciprocal (a value of almost 3) have an about 3.5 percentage points lower probability of searching for a job compared to individuals who are not very negatively reciprocal (value of 2, which is the mean value). Somewhat negatively reciprocal individuals moreover reduce hours spent on job search by 7.6 percent compared to individuals with average negative reciprocity (see Column (2) of Table 5).

Our central result is that negative reciprocity drives the heterogeneity in how unemployed individuals react to stricter regimes. An increase in the strictness of the regime reduces search effort more for negatively reciprocal individuals. If the share of integration agreements enforced increases by one standard deviation (from a mean of 6.6 percent to a share of about 17 percent per team), an individual who is one standard deviation more negatively reciprocal than the average reduces her probability of job search by 2.5 percentage points. Similarly, when increasing the mean strictness of the regime by one standard deviation, an unemployed individual who is somewhat negatively reciprocal reacts with a 6.3 percent reduction in the number of hours spent on job search.

The results allow us to draw the following conclusion: An individual who exhibits average negative reciprocity does not react with an increase in her search effort to stricter regimes. Moreover, the more negatively reciprocal an unemployed individual is, the more she reduces her search effort in response to "sticks" policies. Note that about 35 percent of individuals are somewhat or even more negatively reciprocal. Thus, the group of individuals who potentially reacts negatively to sticks treatments is not negligible.

Our findings imply that also unemployed individuals react to a climate of control and punishment with reduced search effort. Yet, while for employed individuals reduced effort in response to a wage cut harms the firm's output, the consequences for unemployed individuals may be even worse. It suggest that depending on their negative reciprocity "sticks" policies may push unemployed individuals even further away from the labor market.

5.3 Robustness checks

We perform a number of robustness checks. First, we exclude the top 1 percent of strictest regimes to see whether our effects are driven by these teams. This reduces the average share of integration agreements enforced somewhat to about 5.8 percent (standard deviation is now 7.7 percent) on average but barely changes negative reciprocity. Table 6 presents the results. The main effect for the share of integration agreement enforced remains close to zero and insignificant for all measures of search effort. The main effect for negative reciprocity become somewhat larger but remain qualitatively the same as in Table 5. An increase in negative reciprocity by one standard deviation reduces the unemployed's probability of searching for a job by about 3.9 percentage points, and reduces the number of hours she searches by 8.6 percent. The estimated coefficients for the interaction terms are almost double the size of those obtained in Table 5. For job search, this implies that an increase in the strictness of the regime by 7.7 percent at the mean is associated with an 4.6 percentage points lower likelihood of searching for a job when negative reciprocity is by 0.8 points higher than on average. At the same time, this increases the hours spent for job search by 11.7 percent. The estimated effects for the interaction terms suggest that negative reciprocity leads to a more heterogeneous response in the reaction to the strictness of the regimes when the top 1 percent of strictest regimes is excluded. Thus, our main results are not driven by the strictest regimes.

For a second robustness check, we use an alternative definition of the strictness of the regime. Instead of defining strictness as the share of integration agreements per office day, we use the share of integration agreements enforced by an administrative act per 100 clients at the team level. With an average share of 4.3 percent and a standard deviation of 6.8 percent, these numbers are somewhat smaller than for the measure we use in the main specification (see Table 1). The estimated coefficients of interest are provided in Table 7. One can see that they are almost identical in direction and magnitude to those in Table 5. Given mean and standard deviation of the strictness measure, this implies that individuals react relatively stronger. The estimated interaction term effect suggests that a somewhat negatively reciprocal individual reacts to an increase in the strictness of the regime by one standard deviation with a reduction in the probability of searching for a job. The results indicate that our main specification is robust to alternative definition of the strictness of the regime.

In line with the growing interest in gender differences in preferences (Croson and Gneezy, 2009; Falk and Hermle, 2018), we investigate whether the results in Table 5 are driven by one gender only. To this end, we stratify the sample by gender. We first consider the group of male

unemployed individuals.[14] The upper panel of Table 8 presents the results from estimating Equation (1) for men. The estimated coefficients for a change in the strictness of the regime by one standard deviation at the mean value of negative reciprocity are close to zero and insignificant. Increasing negative reciprocity by one standard deviation also tends to reduce effort invested in job search. Although the magnitude is similar to those in Table 5, only the coefficient for log hours spent for job search is statistically significant at the 10 percent level. In contrast, all estimated coefficients for the interaction between negative reciprocity and the strictness of the regime are negative and statistically significant on the 1 percent level. For job search, a somewhat more negative reciprocal individual has a 4 percentage point lower probability to search for a job in response to an increase in the strictness of the regime by one standard deviation than an average individual.

The lower panel of Table 8 displays the regression results for women. There are two striking differences compared to the results for men. First, the estimated main effect for the strictness of the regimes is positive and statistically significant on the 5 percent level for job search and the number of search activities. An increase in the average share of integration agreements enforced by one standard deviation, leads to a 7.2 percentage point higher probability of job search for average negatively reciprocal women. This suggests that women exert more effort in job search in reaction to a stricter regime. Second, the estimated coefficients of the interaction terms are small and not significantly different from zero. Accordingly, negative reciprocity does not play a major role in driving heterogeneous reactions stricter regimes among unemployed women.

6 Conclusion

By investigating the heterogeneous response of long-term unemployed individuals to stricter caseworker regimes, we show that individuals reduce their effort to search for a job when negative incentives are imposed even when they are only somewhat negatively reciprocal. This suggests that for a non-negligible group of long-term unemployed individuals, "sticks" treatments are ineffective. Instead of exerting more effort to find a job, such policies induces to the opposite reaction from the one intended, even among individuals who exhibit only mildly negative reciprocal behavior. It also implies that the effects of negative reciprocity on effort and productivity that have been documented for employed individuals carry over to the group of unemployed individuals. Compared to employed individuals, the additional risk for long-term unemployed is that such "sticks" policies detach them not only even more from

[14]For them, average negative reciprocity is 1.95 (standard deviation 0.75), and the average share of integration agreements enforced through an administrative act is 6.10 (standard deviation 9.01).

the labor market but also fosters social exclusion. Thus, it is questionable whether stricter regimes are effective when the ultimate goal is to bring individuals back to the labor market. In contrast, they may lock-in negatively reciprocal individuals in long-term unemployment.

As a policy implication one may consider tailoring active labor market policies to the personalities of the unemployed. However, before such a road could be pursued, more research is needed to understand how differences in social preferences and personality interact with incentives for unemployed individuals. For example, one issue we leave for future research is whether negative incentives have similar effects on individuals who are short-term unemployed. We also find striking differences by gender. In particular, the reduction in effort invested in job search in response to stricter regimes is driven by males.

References

Akerlof, G.A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97(4), 543–569.

Akerlof, G.A. (1984). Gift exchange and efficiency-wage theory: Four views. *American Economic Review*, 74(2), 79–83.

Arni, P., van den Berg, G.J., and Lalive, R. (2015). Treatment versus regime effects of carrots and sticks. IZA discussion paper series, No. 9457.

Baethge-Kinsky, V., Bartelheimer, P., Henke, J., Wolf, A., Land, R., Willisch, A., and Kupka, P. (2007). Neue soziale Dienstleistungen nach SGB II. *IAB Forschungsbericht*, 15.

Bartelheimer, P. (2008). Wie man an seiner Eingliederung mitwirkt. *Zeitschrift für Sozialreform*, 54(1), 11–36.

Behncke, S., Frölich, M., and Lechner, M. (2010a). A caseworker like me – does the similarity between the unemployed and their caseworkers increase job placements? *The Economic Journal*, 120(549), 1430–1459.

Behncke, S., Frölich, M., and Lechner, M. (2010b). Unemployed and their caseworkers: Should they be friends or foes? *The Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 173(1), 67–92.

Bellemare, C. and Shearer, B. (2009). Gift giving and worker productivity: Evidence from a firm-level experiment. *Games and Economic Behavior*, 67(1), 233–244.

Bender, G., Bieber, D., Hielscher, V., Marschall, J., Ochs, P., and Vaut, S. (2006). *Organisatorischer Umbau der Bundesagentur für Arbeit: Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission*. Institut für Sozialforschung und Sozialwirtschaft e.V., Saarbrücken.

Bewley, T. (1994). *Why wages don't fall during a recession*. Harvard university press, Cambridge.

Bewley, T. (1995). A depressed labor market as explained by participants. *American Economic Review*, 85(2), 250–254.

Black, D.A., J.A. Smith, M.C. Berger, and B.J. Noel (2003). Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System, *American Economic Review*, 93, 1313–1327.

Boockmann, B., C. Osiander, M. Stops, and H. Verbeek (2013). Effekte von Vermittlerhandeln und Vermittlerstrategien im SGB II und SGB III (Pilotstudie), *IAB-Forschungsbericht*, 7/2013.

Brown, M., Falk, A., and Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, 72(3), 747–780.

Camerer, C. and Thaler, R.H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, 9(2), 209–218.

Carpenter, J. and Seki, E. (2005). Competitive work environments and social preferences: field experimental evidence from a japanese fishing community. *Contributions in Economic Analysis & Policy*, 5(2).

Crépon, B. and van den Berg, G.J. (2016). Active labor market policies. *Annual Review of Economics*, 8, 521–546.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47, 448–474.

Dauth, W., Dorner, M., and Blien, U. (2013). Neukonzeption der Typisierung im SGB-II-Bereich: Vorgehensweise und Ergebnisse. *IAB Forschungsbericht*, 11.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo reciprocans: survey evidence on behavioral outcomes. *The Economic Journal*, 119(536), 592–612.

Eichhorst, W., Grienberger-Zingerle, M., and Konle-Seidl, R. (2010). Activating labor market and social policies in Germany: From status protection to basic income support. *German Policy Studies*, 6(1), 65–105.

Englmaier, F. and Leider, S. (2012). Contractual and organizational structure with reciprocal agents. *American Economic Journal: Microeconomics*, 4(2), 146–183.

Fahr, R. and Sunde, U. (2009). Did the Hartz reforms speed up the matching process? A macro evaluation using empirical matching functions. *German Economic Review*, 10(3), 284–316.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293–315.

Falk, A. and Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412).

Fehr, E. and Gächter, S. (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, 42(3), 845–859.

Fehr, E. and Schmidt, K.M. (2006). The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In S.-C. Kolm and J.M. Ythier, editors, *Handbook of the Economics of Giving, Altruism and Reciprocity*, volume 1, chapter 8. Elsevier.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, 108(2), 437–459.

Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, pages 833–860.

Gächter, S. and Fehr, E. (2008). Reciprocity and contract enforcement. In C.R. Plott and V.L. Smith, editors, *Handbook of Experimental Economics Results*, volume 1, chapter 37,

pages 319–324.

Gilchrist, D., Luca, M., and Malhotra, D. (2016). When 3+1>4: Gift structure and reciprocity in the field. *Management Science*, 69, 2639–2650.

Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24(3), 153–172.

Gneezy, U. and List, J. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiment. *Econometrica*, 75, 1365–1384.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.

Gutiérrez-Roig, M., Gracia-Lázaro, C., Perelló, J., Moreno, Y., and Sánchez, A. (2014). Transition from reciprocal cooperation to persistent behaviour in social dilemmas at the end of adolescence. *Nature Communications*, 5, 4362.

Jacobi, L. and Kluve, J. (2007). Before and after the Hartz reforms: The performance of active labour market policy in Germany. *Zeitschrift fuer Arbeitsmarktforschung*, 40(1), 45–64.

Kaltenborn, B., Christe, G., Kupka, P., Schröder, H., Schütz, H., Steinwede, J., and Wielage, N. (2011). *Vermittlung und Beratung in der Praxis. Eine Analyse von Dienstleistungsprozessen am Arbeitsmarkt*. Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Krebs, T. and Scheffell, M. (2013). Macroeconomic evaluation of labor market reform in Germany. *IMF Economic Review*, 61, 664–701.

Kube, S., Maréchal, M.A., and Puppe, C. (2013). Do wage cuts damage work morale? Evidence from a natural field experiment. *Journal of the European Economic Association*, 11(4), 853–870.

Lalive, R., van Ours, J.C., and Zweimüller, J. (2005). The Effect of Benefit Sanctions on the Duration of Unemployment. *Journal of the European Economic Association*, 3(6), 1–32.

Montizaan, R., de Grip, A., Cörvers, F., and Dohmen, T. (2016). The impact of negatively reciprocal inclinations on worker behavior: Evidence from a retrenchment of pension rights. *Management Science*, 62(3), 668–681.

OECD (2013). Activation strategies for stronger and more inclusive labour markets in G20 countries: Key policy challenges and good practices. Report requested by the G20 Task Force on Employment.

Perugini, M., Gallucci, M., Presaghi, F., and Ercolani, A. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17, 251–283.

Ridder, G. (1984). The distribution of single-spell duration data. In G.R. Neumann and N. Westergård-Nielsen, editors, *Studies in Labor Market Analysis*, Springer Verlag, Berlin.

Schütz, H., P. Kupka, S. Koch, and B. Kaltenborn (2011). Eingliederungsvereinbarungen in der Praxis: Reformziele noch nicht erreicht, *IAB−Kurzbericht* 18/2011.

Trappmann, M., Bähr, S., Beste, J., Eberl, A., Frodermann, C., Gundert, S., et al. (2019). Data Resource Profile: Panel Study Labour Market and Social Security (PASS). *International Journal of Epidemiology*, 48(5), 1411–1411g.

van den Berg, G.J. (2001). Duration models: Specification, identification, and multiple durations. In J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

van den Berg, G.J. and Vikström, J. (2014). Monitoring job offer decisions, punishments, exit to work, and job quality. *Scandinavian Journal of Economics*, 116(2), 284–334.

van den Berg, G.J., Hofmann, B., Stephan, G., and Uhlendorff, A. (2014a). Eingliederungsvereinbarungen: Vermittlungsfachkräfte halten mehr Spielraum für sinnvoll. *IAB Kurzbericht 22/2014*.

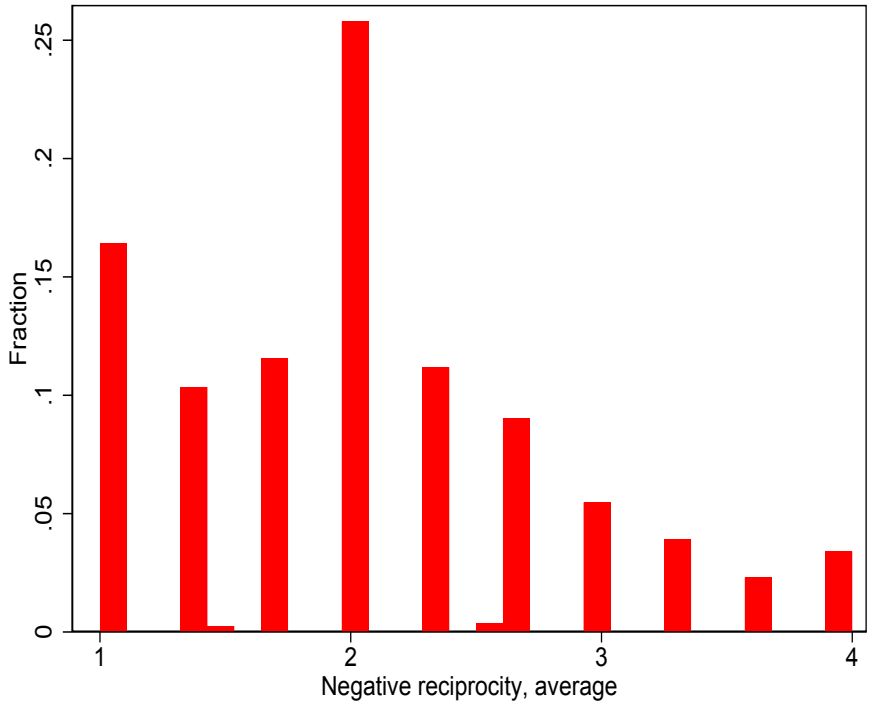
van den Berg, G.J., Hofmann, B., Stephan, G., and Uhlendorff, A. (2014b). Was Vermittlungsfachkräfte von Eingliederungsvereinbarungen halten – Befragungsergebnisse aus einem Modellprojekt. *IAB Forschungsbericht*, 11.

van den Berg, G.J., Uhlendorff, A., and Wolff, J. (2017). Wirkungen von Sanktionen für junge ALG-II-Bezieher: Schnellere Arbeitsaufnahme, aber auch Nebenwirkungen. *IAB-Kurzbericht, 05/2017*.

van der Klaauw, B. and van Ours, J. (2013). Carrot and stick: how reemployment bonuses and benefit sanctions affect exit rates from welfare. *Journal of Applied Econometrics*, 28, 275–296.

Figures

Figure 1: Average negative reciprocity, analytic sample



Tables

Table 1: Summary statistics for main variables, outcomes and covariates, analytic sample

	N	mean	sd	Min	Max
main variables of interest					
negative reciprocity	822	2.040	0.775	1	4
agreement enforced total	822	3.439	5.571	0	46.166
agreement enforced/office days	822	6.555	10.580	0	82.102
agreement enforced/100 clients	822	4.252	6.763	0	45.786
outcomes					
search motivation: high	822	0.560	0.497	0	1
log(hrs spent on job search)	822	1.054	1.008	0	4.331
job search: yes	822	0.629	0.483	0	1
number search activities	822	2.180	1.985	0	7
willingness concede: yes	822	0.564	0.496	0	1
individual characteristics					
male	822	0.546	0.498	0	1
age	822	45.107	11.927	17	64
age quadratic	822	21767	10276	289	4096
number of children	822	0.600	1.054	0	6
unemployment duration	822	47.906	39.330	0	127
migrant	822	0.315	0.465	0	1
migrant: missing	822	0.032	0.175	0	1
highest degree: none	822	0.071	0.256	0	1
highest degree: basic school	822	0.403	0.491	0	1
highest degree: middle school	822	0.288	0.453	0	1
highest degree: university	822	0.215	0.411	0	1
highest degree: other	822	0.023	0.150	0	1
HH type: single person	822	0.473	0.499	0	1
HH type: couple w/0 children	822	0.114	0.318	0	1
HH type: single parent	822	0.206	0.404	0	1
HH type: couple with children	822	0.176	0.381	0	1
HH type: other	822	0.027	0.161	0	1
HH type: missing	822	0.004	0.060	0	1
disability	822	0.136	0.343	0	1
mental health problems	822	0.224	0.417	0	1
physical health problems	822	0.307	0.461	0	1
team-level controls					
share clients male	822	0.504	0.051	0.076	0.726
share clients below age 25	822	0.077	0.169	0	0.944
share clients above age 55	822	0.206	0.130	0	0.865
share clients not German	822	0.245	0.138	0.014	0.600
share clients single parents	822	0.113	0.042	0.012	0.510
share clients market profile	822	0.173	0.119	0.003	0.927
share clients complex profile	822	0.519	0.141	0.005	0.948
share clients other profiles	822	0.254	0.098	0.008	0.920
share clients no profiles classified yet	822	0.054	0.046	0	0.463

Table 2: Correlations (panel A) and summary statistics (panel B) for using an administrative act to enforce an integration agreement, 2012–2015

Panel A: raw correlations across years

share integration agreements enforced per year, N=375	2015	2014	2013	2012
2015	1			
2014	0.9556	1		
2013	0.8909	0.9470	1	
2012	0.8244	0.8687	0.9167	1

Panel B: Summary statistics per year

share integration agreements enforced per year	mean	sd	min	max
2015 (N= 497)	3.226	4.867	0	46.059
2014 (N=478)	3.110	4.888	0	44.241
2013 (N=443)	3.236	5.391	0	49.440
2012 (N=376)	2.834	4.698	0	39.731

Table 3: Results from an OLS regression of negative reciprocity on the share of integration agreements enforced

	(1)	(2)	(3)
VARIABLES		negative reciprocity	
agreement enforced/office days	-0.001	-0.003	-0.004
	(0.004)	(0.003)	(0.003)
constant	2.049***	2.012***	3.416***
	(0.037)	(0.397)	(0.865)
observations	822	822	822
individual controls	No	Yes	Yes
team-level controls	No	No	Yes
job center type FE	No	No	Yes

$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$; standard errors are clustered at the team level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all contact days of a team; controls on individual level: dummies for highest educational degree, gender, age, age quadratic, dummies for household types, number of kids in household, dummy for ethnic German, dummy for poor physical health, dummy for disability, missing value flag for household type and migration; controls at team level: fraction non-German clients, fraction clients below age of 25, fraction clients single mum/dad, fraction female clients, fraction of unemployed in each profile type; job center type FE.

Table 4: OLS regression of the share of integration agreements enforced per team on team-level characteristics

	(1)	(2)
VARIABLES	agreement enforced	agreement enforced
share clients male	0.945	0.903
	(1.818)	(1.626)
share clients below age 25 25	0.092	0.164
	(0.312)	(0.307)
share clients above age 55	-0.731	-0.995***
	(0.493)	(0.339)
share clients not German	0.377	0.847
	(0.607)	(0.694)
share clients single parents	1.259	-0.212
	(1.789)	(1.244)
share clients complex profile	0.186	0.351
	(0.544)	(0.502)
share clients other profiles	-0.105	-0.391
	(0.442)	(0.359)
share clients no profile classified yet	-0.355	-0.678
	(0.905)	(0.792)
Constant	-0.649	2.388
	(1.383)	(1.700)
job center type FE	no	yes
Observations	477	477
R-squared	0.019	0.202

$^{***}p < 0.01,^{**}\ p < 0.05,^{*}\ p < 0.1$; standard errors are clustered on the job center level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all contact days of a team; job center FE: classification of job centers into 15 categories according to selected characteristics of the local labor market such as rural/urban area, share migrants, share high educated, high job density, industry classification. For details see Dauth *et al.* (2013).

Table 5: Results from an OLS regression of search outcomes on severity of team-level regime and negative reciprocity

VARIABLES	(1) jobsearch yes	(2) log(hrs job search)	(3) search motivation	(4) willingness concede	(5) # search activities
std agr. enforced/cont. days	0.017	0.018	0.021	0.003	0.029
	(0.018)	(0.041)	(0.018)	(0.021)	(0.063)
std negrec	-0.035**	-0.076**	-0.027	-0.029*	-0.091
	(0.017)	(0.035)	(0.017)	(0.016)	(0.069)
std agr. enforced/cont. days × std negrec	-0.025*	-0.063**	-0.033***	-0.025**	-0.143***
	(0.013)	(0.026)	(0.012)	(0.012)	(0.048)
constant	1.126**	2.756**	1.617***	1.128**	2.933
	(0.520)	(1.104)	(0.527)	(0.544)	(2.093)
observations	822	822	822	822	822

$^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$; standard errors are clustered at the team level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all contact days of a team; controls on individual level: dummies for highest educational degree, gender, age, age quadratic, dummies for household types, number of kids in household, dummy for ethnic German, dummy for poor physical health, dummy for disability, missing value flag for household type and migration; controls at team level: fraction non-German clients, fraction clients below age of 25, fraction clients single mum/dad, fraction female clients, fraction of unemployed in each profile type, job center type FE.

Table 6: Results from an OLS regression of search outcomes on severity of team-level regime and negative reciprocity, sample exclude top 1% of strictest teams

VARIABLES	(1) jobsearch: yes	(2) log(hrs job search)	(3) search motivation	(4) willingness concede	(5) # search activities
std agr. enforced/cont. days	0.013	0.001	0.014	-0.013	0.037
	(0.022)	(0.048)	(0.021)	(0.027)	(0.086)
std negrec	-0.039**	-0.086**	-0.032*	-0.032*	-0.102
	(0.017)	(0.035)	(0.018)	(0.017)	(0.070)
std agr. enforced/cont. days × std negrec	-0.046*	-0.117***	-0.058**	-0.040	-0.202**
	(0.027)	(0.041)	(0.026)	(0.027)	(0.095)
constant	0.934*	2.362**	1.461***	0.924	2.163
	(0.551)	(1.152)	(0.560)	(0.565)	(2.306)
observations	813	813	813	813	813

$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$; standard errors are clustered at the team level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all contact days leaving out the highest percentile; controls on individual level: dummies for highest educational degree, gender, age, age quadratic, dummies for household types, number of kids in household, dummy for ethnic German, dummy for poor physical health, dummy for disability, missing value flag for household type and migration; controls at team level: fraction non-German clients, fraction clients below age of 25, fraction clients single mum/dad, fraction female clients, fraction of unemployed in each profile type, job center type dummies

Table 7: Results from an OLS regression of search outcomes on severity of team-level regime and negative reciprocity, alternative definition of severity of team-level regime

VARIABLES	(1) jobsearch: yes	(2) log(hrs job search)	(3) search motivation	(4) willingness concede	(5) # search activities
std agr. enforced/100 clients	0.013	0.002	0.020	-0.002	0.026
	(0.018)	(0.040)	(0.018)	(0.020)	(0.063)
std negrec	-0.035**	-0.077**	-0.027	-0.029*	-0.093
	(0.017)	(0.035)	(0.017)	(0.016)	(0.069)
std agr. enforced/100 clients × std negrec	-0.030**	-0.065**	-0.037***	-0.029**	-0.158***
	(0.015)	(0.026)	(0.014)	(0.013)	(0.051)
constant	1.145**	2.789**	1.639***	1.133**	2.967
	(0.519)	(1.105)	(0.525)	(0.539)	(2.088)
observations	822	822	822	822	822

$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$; standard errors are clustered at the team level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all unemployed in the same team; controls on individual level: dummies for highest educational degree, gender, age, age quadratic, dummies for household types, number of kids in household, dummy for ethnic German, dummy for poor physical health, dummy for disability, missing value flag for household type and migration; controls at team level: fraction non-German clients, fraction clients below age of 25, fraction clients single mum/dad, fraction female clients, fraction of unemployed in each profile type, job center type dummies

Table 8: Results from an OLS regression of search outcomes on severity of team-level regime and negative reciprocity, by gender of the unemployed individual

VARIABLES	(1) jobsearch: yes	(2) log(hrs job search)	(3) search motivation	(4) willingness concede	(5) # search activities
			Men, N=449		
std agr. enforced/cont. days	-0.007	-0.015	0.001	-0.010	-0.119
	(0.022)	(0.056)	(0.022)	(0.024)	(0.076)
std negrec	-0.031	-0.098*	-0.020	-0.033	-0.072
	(0.023)	(0.050)	(0.025)	(0.023)	(0.096)
std agr. enforced/cont. days × std negrec	-0.040***	-0.086***	-0.045***	-0.037***	-0.184***
	(0.012)	(0.019)	(0.012)	(0.012)	(0.037)
Constant	0.377	2.483	1.120	0.293	-0.606
	(0.755)	(1.723)	(0.796)	(0.789)	(3.088)
			Women, N=373		
std agreement enforced	0.072**	0.067	0.042	0.012	0.318**
	(0.031)	(0.059)	(0.033)	(0.041)	(0.125)
std negrec	-0.032	-0.022	-0.028	-0.028	-0.098
	(0.028)	(0.057)	(0.029)	(0.028)	(0.106)
std agreement enforced × std negrec	0.021	-0.016	-0.016	-0.009	0.021
	(0.029)	(0.063)	(0.033)	(0.028)	(0.131)
Constant	1.652**	2.787*	1.955***	1.776**	6.234**
	(0.677)	(1.431)	(0.699)	(0.749)	(2.605)

$^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$; standard errors are clustered at the team level; agreement enforced is defined as the share of integration agreements that were enforced through an administrative act among all unemployed in the same team; controls on individual level: dummies for highest educational degree, gender, age, age quadratic, dummies for household types, number of kids in household, dummy for ethnic German, dummy for poor physical health, dummy for disability, missing value flag for household type and migration; controls at team level: fraction non-German clients, fraction clients below age of 25, fraction clients single mum/dad, fraction female clients, fraction of unemployed in each profile type, job center type dummies