Political Science Research and Methods (2024), **12**, 783–798
doi:10.1017/psrm.2024.1

ORIGINAL ARTICLE

Attitudes toward automation and the demand for policies addressing job loss: the effects of information about trade-offs

Beatrice Magistro[1], Peter Loewen[2], Bart Bonikowski[3], Sophie Borwein[4] and Blake Lee-Whiting[5]

[1]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA, [2]Munk School of Global Affairs and Public Policy, University of Toronto, Toronto, ON, Canada, [3]Department of Sociology, New York University, New York City, NY, USA, [4]Department of Political Science, University of British Columbia, Vancouver, BC, Canada and [5]Department of Political Science, University of Toronto, Toronto, ON, Canada
Corresponding author: Beatrice Magistro; Email: magistro@caltech.edu

(Received 24 February 2023; revised 21 June 2023; accepted 13 September 2023; first published online 15 February 2024)

Abstract
Does providing information about the costs and benefits of automation affect the perceived fairness of a firm's decision to automate or support for government policies addressing automation's labor market consequences? To answer these questions, we use data from vignette and conjoint experiments across four advanced economies (Australia, Canada, the UK, and the US). Our results show that despite people's relatively fixed policy preferences, their evaluation of the fairness of automation—and therefore potentially the issue's political salience—is sensitive to information about its trade-offs, especially information about price changes attributable to automated labor. This suggests that the political impact of automation may depend on how it is framed by the media and political actors.

Keywords: automation; conjoint experiment; costs and benefits; fairness; methodology; policy preferences; political economy; survey; technological change; trade-offs

1. Introduction

Advanced industrial economies are experiencing major changes, including labor dislocation, as a result of advances in automation and artificial intelligence (AI). While automation is likely to lead to aggregate efficiency gains, it also has distributional consequences, as it will cause some workers to lose their jobs while creating new employment opportunities for others (Autor *et al.*, 2003; Autor and Dorn, 2013; Bessen, 2019; Aghion *et al.*, 2020, 2021). As a result of the heterogeneity of these labor market disruptions, there is a growing literature on the political consequences of automation and AI (Frey and Osborne, 2017; Im *et al.*, 2019; Anelli *et al.*, 2019; Di Tella and Rodrik, 2020; Gallego *et al.*, 2022). Much of this research finds that workplace automation matters for voting behavior (Anelli *et al.*, 2019; Im *et al.*, 2019; Kurer, 2020). People adversely affected by technological change are more likely to vote against the political establishment; whether they lean left or right depends on how they understand technological threat (Borwein *et al.*, 2024a). The findings in the literature, however, are less consistent when it comes to the effects of automation-related risk exposure on workers' policy preferences: automation risk is not always related to preferences for more redistribution or for policies protecting jobs (Gallego and Kurer, 2022).

Much of the existing literature draws upon standard political economy models, according to which labor market risks affect economic interest, which in turn affects political preferences and voting behavior. But are people aware of the trade-offs of automation for labor markets, economic growth, and the prices of goods and services? If so, how do they evaluate those trade-offs? These questions are particularly important because the political consequences of different economic shocks vary based on how they are perceived by voters and how they are interpreted by political actors.

There are reasons to believe that voters may not be aware of the potential effects of automation. First, although technological change has been the largest determinant of job polarization, contemporary political parties have not consistently or widely mobilized voters against technological change; instead they have turned to other explanations for recent structural changes in labor markets, like trade and immigration, that tend to fuel out-group resentments (Mutz, 2021; Wu, 2021; Gallego and Kurer, 2022).[1]

At the same time, despite the evident economic benefits stemming from technological change, automation is often discussed negatively by the media and the scholarly literature, which tend to frame it as a cause of job loss rather than job creation (Frey and Osborne, 2017; Im *et al.*, 2019; Anelli *et al.*, 2019). Existing experimental research on attitudes toward automation similarly relies on treatments that describe only the costs of technological change, stressing the number of jobs that may be lost (Di Tella and Rodrik, 2020; Jeffrey and Matakos, 2021; Jeffrey, 2021; Mutz, 2021; Werfel *et al.*, 2022), or that communicate trade-offs only in vague terms, without providing concrete information about the magnitude of the positive and negative changes associated with automation (Gallego *et al.*, 2021). An emphasis on costs alone, however, may increase the perceived unfairness of automation, which may in turn heighten popular opposition against it, as has likely been the case for trade (Rodrik, 2018).

In this paper, we examine opinions toward automation more broadly. In particular, we ask: do people perceive a firm's decision to automate to be fairer if information on both the costs and benefits is disclosed, instead of the costs alone? Does providing concrete information on the trade-offs of automation, rather than only its costs, affect support for different government policies targeting automation's negative labor market consequences? To examine these questions, in our experimental treatments, we provide information not only about potential job and wage losses, but also about potential job and wage gains, as well as potential consumer price changes resulting from the introduction of new technology.

We use a combination of vignette and conjoint experiments across four advanced, liberal market economies (Australia, Canada, the UK, and the US), with a total sample of about 8000 individuals, to answer these questions. Respondents are randomly assigned into one of three conditions: a costs-only news article treatment, a generic trade-off information treatment, or a specific trade-off information conjoint treatment. In all treatment groups individuals read about a firm introducing a new computer-based productivity-improving technology. In the news article treatment individuals are only exposed to the costs of automation, i.e., job and wage losses, and are not given information about the benefits. Respondents in the generic trade-off information group read that the firm's decision will lead to some job gains, some job losses, and possibly lower prices, but are not provided with estimates of the magnitude of these effects. In the specific trade-off information conjoint treatment each individual sees an iteration of four random tables, each with varying quantified costs and benefits of automation (in terms of final products' price changes and changes in the number of employed workers and their wages). Finally, after the respective vignettes, individuals in all groups are asked how much they perceive the firm's decision to automate as fair and what government policies in response to automation they favor.

[1]By job polarization we mean the relative growth of employment in high-skill jobs and low-skill jobs amid the concurrent decline in middle-skill jobs (Autor and Dorn, 2013; Goos *et al.*, 2014).

Our analysis proceeds in two steps. First, we examine results from the between-subjects experiment with three groups: the news article treatment, the generic information treatment, and the specific information conjoint (for the latter we average each respondent's answers across the four trade-off tables they were shown). We then examine the perceived fairness of the firm's decision to automate and respondents' attitudes toward potential policy responses (i.e., social spending, basic income, job guarantees, unskilled immigration restrictions, skilled immigration restrictions, trade restrictions, retraining workers, and taxing automation).

Second, we use conjoint analysis to examine respondents' attitudes toward different trade-off scenarios. This analysis tells us which attributes causally increase or decrease, on average, the perceived fairness of the firm's decision to automate when other attributes are held constant. Attributes include type of product, increases or decreases in the number of high-skilled workers, number of low-skilled workers, wages of high-skilled workers, wages of low-skilled workers, and product price.

We find that relative to the costs-only news treatment, people in the generic information group (i.e., those exposed to the trade-offs of automation framed in vague terms) perceive the firm's decision to automate as fairer. Respondents in the specific information group, who were shown numeric estimates of the costs and benefits of automation, also perceive the decision to automate as fairer than people in the costs-only news condition, but this increase is not as large as for those in the generic information group. This suggests that people exposed to the generic trade-off information may envision the net gains of automation to be larger relative to respondents in the specific information treatment group.

When it comes to support for policies addressing the consequences of automation, however, there are no discernible differences across the three treatment groups, suggesting that policy preferences are relatively sticky. Consistent with other studies (Di Tella and Rodrik, 2020; Gallego *et al.*, 2023), respondents in all three groups favor retraining individuals affected by automation above any other policy option. The least popular policy responses are restricting skilled immigration and trade.

Finally, our conjoint experiments present subjects with information on consumer price changes and job and wage losses or gains for both high- and low-skilled employees. Our results show that respondents are most sensitive to price changes. When prices decrease by 20–50 percent as a result of the new technology, respondents are much more likely to perceive automation as fair than when there is no price change. Similarly, when fewer (more) jobs are lost (gained), they are more likely to see automation as fair than when more (fewer) jobs are lost (gained). They are also sensitive to decreases in the wages of workers affected by automation, but not to wage increases. Respondents' attitudes do not appear to be affected by the type of firm or product in question, whether the treatment involves an auto company producing cars, an aviation firm manufacturing planes, an electronics company making smartphones, or a pharmaceutical company producing vaccines.

Our research makes several contributions. First, we focus on the relative costs and benefits of automation, in contrast to the majority of existing studies and media reports, which tend to provide pessimistic accounts focused solely on automation's negative consequences. We begin with an information treatment that emphasizes the costs of automation, in line with the dominant scholarly and journalistic accounts, and compare this baseline scenario to two information treatments that respectively emphasize generic and specific trade-offs of automation (using either vague or quantified costs and benefits) to see how these primes affect the perceived fairness of a firm's decision to automate and the demand for policies addressing job loss. Importantly, we find that relative to a costs-only message, information on trade-offs can significantly increase the perceived fairness of automation. This is especially true of generic information; more specific information leads to smaller increases in perceived fairness, on average, than vague information, but this varies based on the relative magnitude of gains and losses from automation, as highlighted by our conjoint analysis. Second, people's policy preferences are relatively stable—none

of the information treatments shift respondents' preferences across policy responses, as they are consistently more supportive of retraining workers relative to other policy options. Finally, ours is the first experimental study to include consumer price changes among the potential benefits or costs of automation. We find that this is the feature to which respondents are by far most sensitive. While respondents also react to changes in employment and, to a lesser degree, in wages, when prices decrease by 20 or 50 percent they are much more likely to perceive the firm's decision to automate as fair. These findings suggest that the political consequences of automation may differ based on whether political actors decide to politicize the issue and what information they provide to voters.

2. The political economy of automation

Labor markets may be profoundly reshaped by technological change, in particular by automation and AI. While many analysts have predicted that these developments will lead to large-scale job displacement as new technologies replace labor, others have argued that these fears are unwarranted in light of historical evidence about technological change. Past technological revolutions that yielded the largest gains in growth and prosperity—the steam engine in the early 1800s and electricity in the 1920s—did not lead to mass unemployment as some had anticipated (Aghion *et al.*, 2021). The literature has thus shifted from a predominantly negative view of automation as a job destroyer to a more positive outlook that sees its economic and social consequences as more benign (Bessen, 2019; Aghion *et al.*, 2020, 2021). In particular, the latter perspective highlights the direct impact of automation on productivity: companies that implement automation experience a boost in productivity, which enables them to reduce their quality-adjusted prices. This reduction in prices leads to a rise in demand for their products, ultimately resulting in increased employment within these firms.

In addition to the aggregate effects of automation on employment, scholars have also considered its distributional consequences. In particular, recent research shows that job losses have been concentrated in occupations involving routine tasks, as opposed to jobs requiring human interaction and higher education, which are least at risk (Autor *et al.*, 2003; Autor and Dorn, 2013). Some studies find evidence of a reallocation of workers between occupations (Humlum, 2019), with labor demand shifting from low-skilled to high-skilled workers. Kurer and Gallego (2019) focus on the aggregate decline in routine work as a result of technological change and find that many routine workers manage to keep their jobs until early retirement, and that employment decline is mostly determined by higher exit rates and lower entry rates rather than layoffs. Despite varying estimates regarding the risk to jobs, there is consensus among experts that automation and AI will persistently alter labor markets (Autor *et al.*, 2020). This transformation will result in some workers losing their jobs to automation, others finding new employment opportunities, and a significant number needing to develop new skills to successfully transition between occupations.

Given these heterogeneous labor market disruptions, scholars are increasingly interested in the political consequences of technological change. Several studies have investigated the effects of technological change on vote choice, whereas others have looked at citizens' attitudes toward automation and their preferred policies to address technological change. Findings from the former suggest that a greater risk of job loss from automation, whether subjective or objective, is related to both left and populist right voting (Frey *et al.*, 2018; Im *et al.*, 2019; Anelli *et al.*, 2019; Borwein *et al.*, 2024a). This literature, however, has mostly focused on left-behind voters, while neglecting the large majority of workers who stand to benefit from innovation. Two exceptions are Gallego *et al.* (2022) and Schöll and Kurer (2024). Gallego *et al.* (2022)'s study of the active labor force in the UK between 1997 and 2017 provides evidence that technological adoption was economically beneficial for workers with middle and high levels of education but produced small negative effects for low-education workers. Furthermore, growth in automation

increased support for the incumbent party and voter turnout among those who benefited from technological change. Similarly, using data from Germany, Schöll and Kurer (2024) show that while manufacturing and routine jobs declined in regions with high robot adoption or ICT investment, this was offset by an increase in jobs in the service sector and cognitive non-routine occupations. This shift in the composition of the workforce had significant political implications, as workers in these new sectors tend to hold progressive policy preferences and support mainstream, pro-system parties. In conclusion, the authors suggest that the positive effect of the changing workforce composition outweighs the potentially negative effects of automation-induced job substitution.

2.1 Perceptions of fairness

Views of organizations' technology-related decisions—such as whether to automate—and governments' policy responses to the negative externalities of such decisions are influenced by people's perceptions of fairness (Kahneman *et al.*, 1986; Alesina and Angeletos, 2005; Alesina and Giuliano, 2011; Rodrik, 2018; Ciccone *et al.*, 2020; Jeffrey, 2021). From the perspective of a firm, while unfair behaviors—such as paying employees lower wages or charging customers higher prices—may result in short-term gains, these can be outweighed by long-term costs, such as reputational penalties and decreased customer and employee satisfaction. This, in turn, can affect revenues and the ability to attract top employees. Studies have shown that customers may switch their patronage and employees may look for other opportunities if they perceive companies' behavior to violate fairness norms (Kahneman *et al.*, 1986; Rubin, 2012). Survey research suggests that people are often willing to forego personal financial gain to ensure fairness, which may constrain profit-seeking behavior and influence market outcomes (Kahneman *et al.*, 1986). In the context of trade, Rodrik (2018) argues that what arouses popular opposition against trade is not inequality per se, but perceived unfairness. For example, citizens express greater opposition to trade with countries that use illegitimate means to gain a competitive edge (e.g., those with a record of serious labor abuses) (Di Tella and Rodrik, 2020). Similarly, Starmans *et al.* (2017) find no evidence that people are actually concerned about economic inequality; rather, what they worry about is the lack of a level economic playing field. When it comes to automation, relative to a control group unexposed to automation or job loss information, Jeffrey (2021) finds that an information treatment emphasizing an automation shock using politicized rhetoric significantly increases the belief that growth in automation-induced inequality would be unfair, and this in turn increases support for redistributive policies in response. Conversely, Ladreit (2022) finds that a treatment conveying information about an automation shock increases respondents' propensity to see firms as justified in automating compared to a control group, but it does not affect preferences for redistribution.

2.2 Policy preferences

In terms of preferences for policies addressing automation, previous research suggests that individuals who perceive a high degree of technological risk tend to advocate for increased social protection from the welfare state (Thewissen and Rueda, 2019; Sacchi *et al.*, 2020). This can take several forms: social transfers and compensatory schemes; social investment, including active labor market policies; and protectionism. Social transfers and compensatory schemes are aimed at directly compensating workers for job or wage losses related to technological change, for instance, through more generous unemployment benefits or less conventional programs, such as universal basic income schemes. The latter are less targeted than unemployment benefits but they are nonetheless a type of redistributive instrument (Martinelli, 2020). Proponents of the social investment welfare state model (Hemerijck, 2018) highlight the importance of investing in human capital development throughout an individual's life to proactively address social risks such as unemployment. One key component of this approach are active labor market policies,

which involve longer-term solutions involving training or retraining opportunities for workers. However, when it comes to certain labor market shocks, such as those arising from globalization, protectionist policies may be favored over compensatory or social investment ones, as they allow workers to preserve their jobs, although at a much larger cost to society. When it comes to technology, workers may favor policies that attempt to prevent or disincentivize the adoption of new technologies (Abbott and Bogenschneider, 2018; Dauth *et al.*, 2021). Citizens may also react to fears of automation and job insecurity by directing blame and punitive attitudes toward groups they deem undesirable; they may, for example, call for governmental measures against other sources of job threats that they attribute to external factors, such as immigration or trade (Wu, 2023).

Focusing specifically on automation threats, Kurer and Häusermann (2011) present evidence suggesting a positive association between automation risk and support for redistribution, though not necessarily for social investment policies. Conversely, Jeffrey (2021), based on a survey experiment conducted in the UK, and Zhang (2022), using comparable US data, do not find evidence that simulated automation shocks increase support for redistribution or other related policies. Dermont and Weisstanner (2020) look at the relationship between technological risk and support for basic income and find no association, while Im (2021) argues that automation risk is positively linked to support for active labor market policies. Di Tella and Rodrik (2020) prime respondents with several types of labor market disruptions and find that both trade shocks and automation are positively associated with the demand for protectionism but not other social transfers (the effects of trade are stronger than those of automation). Other studies provide more mixed findings, including null effects of automation primes on public policy preferences, with considerable variation across social subgroups (Jeffrey, 2021; Ladreit, 2022; Wu, 2023; Zhang, 2022). In summary, these findings indicate mixed results regarding the impact of automation exposure on support for ameliorative policy responses.

3. Theoretical expectations

Inspired by the literature on automation and other labor market shocks, our paper addresses several questions: Do people perceive a firm's decision to automate to be fairer if information about both the costs and benefits of automation is disclosed, relative to information about its costs only? Does providing information on the trade-offs of automation, relative to its costs only, affect support for specific policies addressing automation's negative consequences? Under what conditions are citizens more or less likely to support retraining workers or providing unemployment insurance for displaced workers as opposed to favoring protectionism? Findings from past studies enable us to make some broad predictions.[2]

First, we expect that when automation is presented as generating greater costs than benefits, or when only its costs are discussed, people will perceive it as more unfair. We know that automation tends to be portrayed negatively by the media and the scholarly literature, as it is more often associated with job loss than job creation (Frey and Osborne, 2017; Anelli *et al.*, 2019; Im *et al.*, 2019). A case in point is the prediction by Frey and Osborne (2017) that 47 percent of current jobs could be lost to automation in the next 20 years, which has attracted significant scholarly and media interest (Gallego *et al.*, 2021). Given that concerns about fairness are deeply ingrained in human cognition as a mechanism for sanctioning opportunistic behavior, it is plausible that when individuals are presented with information that only mentions the negative effects of a firm's decision to automate (i.e. job and wage losses), they will be more likely to perceive that decision as unfair compared to respondents presented with a more balanced scenario, particularly

[2]We registered our pre-analysis plan on OSF at https://osf.io/a8x52. We have diverted somewhat from our PAP where we felt that doing so would make the presentation of the data clearer—in those instances, we changed the framing but not the predictions.

when gains from automation are shown to exceed the losses. Seeing the broader societal benefits of automation (job and wage gains and lower prices) may make people perceive any losses as less unfair. People's perception of fairness may subsequently affect their favorability toward firm-level decisions related to automation, as has been shown for several other economic domains, including decisions to support or engage in trade (Rodrik, 2018; Ciccone *et al.*, 2020).

To the degree that people are concerned about automation, which public policies are they most likely to support in response? We expect that this may depend on the information to which respondents are exposed regarding the combination of costs and benefits associated with the new technologies. Existing studies suggest that exogenous economic shocks lead to increased support for general welfare spending (Margalit, 2013), but we know comparatively little about the types of policies that individuals may support under different information scenarios. Some studies suggest that compensatory policies are more likely to be favored than social investment measures, especially in situations of high economic insecurity (Marx, 2014; Neimanns *et al.*, 2018; Han and Kwon, 2020). In general, we anticipate that individuals who perceive technological change as an imminent threat will be more likely to favor protectionist policies over compensatory policies and compensatory policies over social investment policies. We theorize that in scenarios in which only the costs of automation (i.e., job and wage loss) are made salient, people will exhibit more short-term policy preferences, such as protectionism, in an effort to directly save jobs, followed by compensatory attitudes, aimed at achieving basic protection against loss of income in the event of job loss, rather than favoring longer-term solutions such as social investment. Conversely, in scenarios in which people are exposed to both quantified costs and benefits, when both labor market uncertainty and risk are presumably lower, respondents should on average be more likely to support social investment policies (such as re-training) over policies that are compensatory (such as a basic income or unemployment benefits), protect jobs (including against other risks like trade or immigration), or inhibit automation (either directly or through taxation). When benefits are also quantified alongside costs, the perceived risks of automation are likely to be dampened and the costs of protectionist or inhibitory policies (including foregone benefits of automation) should be clearer.

Furthermore, different welfare state systems may give rise to different reactions to labor market shocks (Busemeyer and Tober, 2023), as individuals' policy preferences are likely to be affected by the surrounding institutional environment. People living in residual welfare states with less generous (or entirely lacking) protection schemes, such as the liberal market economies on which we focus in our study, may perceive higher risks and labor market uncertainties stemming from automation and, if so, they may be even more likely than respondents in more generous welfare states to prioritize protectionism over compensation and compensation over social investment (Busemeyer and Tober, 2023). Although we lack comparative data to test this prediction directly, we expect protectionism to be especially salient among our respondents.

Using a conjoint design, we also investigate which attributes, on average, causally increase or decrease the perceived fairness of the firm's decision to automate when other attributes are held constant. We consider the type of product manufactured by the company, the number of high-skilled jobs gained, the number of low-skilled jobs lost, the changes in the wages of high-skilled workers, the changes in the wages of low-skilled workers, and the estimated price change of products manufactured by the company. We are agnostic as to which of the attributes may matter most, but we expect respondents to be sensitive to information about greater costs or benefits for each attribute: for example, as prices decrease as a result of technological innovation, respondents should become more supportive of automation, all else equal. In line with prospect theory (Kahnemann and Tversky, 1979), it is also possible that job and wage losses may be weighted more heavily than job and wage gains.

Most studies on both automation and trade attitudes measure respondents' self-interest based on the characteristics of their employment or income sources, not on the prices of goods and services that they buy (Baker, 2003; Naoi and Kume, 2015). However, at least

theoretically, respondents may exhibit a greater concern for prices compared to job and wage losses and gains. Prices affect everyone as consumers, regardless of their employment status or industry, by directly influencing their purchasing power and economic well-being. As such, price fluctuations are more universally experienced and immediately recognizable in individuals' daily lives. This may make prices especially salient when respondents assess the consequences of automation. In contrast, job losses and gains resulting from automation tend to be concentrated within specific industries or occupations. Individuals outside of the affected industries or occupations mentioned in the specific information treatment may perceive the hypothetical job losses or gains as more distant and less personally relevant to their own circumstances, compared to respondents with direct experience in those industries or occupations. Individuals may also be especially responsive to price changes because this aspect of automation has received less coverage by the media than job and wage losses, making the information more novel. The evidence is mixed on whether people care more about labor market or consumption-related changes in the context of trade (Baker, 2003; Naoi and Kume, 2015; Chatruc *et al.*, 2021), but to our knowledge, this is the first study to examine these trade-offs with respect to automation.[3]

4. Survey experiments

To assess the role of cost-benefit information on the perceived fairness of a firm's decision to automate and on support for a range of policy responses, we fielded surveys in Australia, Canada, the UK, and the US.[4] Participants were recruited from Cint, an online survey sample platform, in March 2022.[5] Cint provides quota samples that closely mirror the marginal distribution of important demographic variables in the population. We specifically instructed Cint to recruit respondents over 18 years old and to under-sample individuals over 65, since we wanted a sample reflective of working-age adults; furthermore, we applied quotas for age, gender, and region. Excluding respondents with questionable IP addresses, duplicate Cint IDs, and fast completion times (below the second percentile), the sample consisted of 8033 respondents: 1955 in Australia, 1972 in Canada, 2031 in the UK, and 1966 in the US, with a median completion time of 15 minutes. We assigned 2/3 of respondents to the specific information conjoint treatment group, 1/6 to the news treatment group, and 1/6 to the generic information group. Further details regarding the sampling method, quotas, and treatment group randomization are available in the online Appendix. Although Cint's samples are not representative of each country's general population, they approximate the population distributions on key variables, while maximizing the internal validity of experimental findings, in line with many prior studies of labor market shocks (Di Tella and Rodrik, 2020; Wu, 2023; Zhang, 2022). Furthermore, previous research using online platforms, such as Amazon's Mechanical Turk and Lucid, has demonstrated that convenience and representative samples can yield comparable experimental results (Berinski *et al.*, 2012; Huff and Tingley, 2015; Mullinix *et al.*, 2015; Coppock and McClellan, 2019).

We estimate all models using ordinary least squares and pooling data from all four countries.[6] We also conducted separate analyses by country but found little variation in attitudes toward

[3]It is possible that these effects may vary by respondents' economic literacy. Magistro (2022) finds that people who correctly compute the costs and benefits of a policy are more likely to support or oppose that policy. We explore this in the online Appendix where we measure the moderation effects of objective knowledge in the conjoint experiment by asking respondents to compute the costs and benefits of the new innovation. We find that people with higher objective or subjective knowledge of automation are more sensitive to price changes and to changes in the number of employed high-skilled workers.

[4]We provide more information on case selection and on the survey design in the online Appendix.

[5]Fielding occurred between 11 and 28 March 2022. For more information about the sample provider see https://www.cint.com/

[6]The exclusion of respondents who failed a simple attention check produces substantively similar results. The attention check item asks respondents to select "Somehow disagree" among the response categories.

automation and policy preferences. Consequently, we only report the country-by-country analyses in the online Appendix. The following sections describe each treatment condition.

4.1 News article condition

In the news article treatment vignette, individuals read about the costs of automation—with reference to job and wage losses—but were not given information about its benefits:[7]

> Assembly and factory jobs are at risk at a manufacturing firm in [Australia, Canada, the UK, the US], as management has decided to introduce a new computer-based productivity-improving technology, which would lower production costs significantly. We interviewed an employee there for 20 years, who said that the technology shock will be devastating: "Up to 150 people will become unemployed and the rest would have to accept lower wages," he added.

After viewing the vignette, respondents were asked how much they agreed or disagreed with the statement [from Borwein *et al.* (2024b)] that "The company's decision to introduce the new technology is fair," using a five-point scale from "strongly disagree" to "strongly agree" (with a sixth "don't know" option).[8] They were then asked how much they agreed or disagreed with each of the following government policies designed to mitigate the negative consequences of automation (again on a five-point scale with a "don't know" option): expanding social spending, implementing a basic income, guaranteeing workers jobs, reducing the number of unskilled and skilled immigrants, restricting trade, reskilling workers, and introducing an automation tax (see the online Appendix for exact phrasing).

4.2 Generic trade-offs condition

Respondents in the generic trade-offs condition were exposed to information about a new innovation that involves both costs and benefits, but were not given any precise estimates of their relative magnitude:

> A manufacturing firm in [Australia, Canada, the UK, the US] decides to introduce a new computer-based productivity-improving technology. As a result of this innovation, production costs will decrease, and the price of the company's final products could also decrease. Furthermore, while some jobs will be gained, others will be lost.

After viewing the vignette, respondents were asked the same questions as in the news article condition about the perceived fairness of the firm's decision to automate and support toward different policy responses.

4.3 Specific trade-offs conjoint condition

Individuals in the conjoint group first saw a pre-treatment prompt:

> A manufacturing firm in [Australia, Canada, the UK, the US] decides to introduce a new computer-based productivity-improving technology. As a result of this innovation, production costs will decrease, and the price of the company's final products could also decrease.

[7]The source of the vignette text is not communicated in the experiment. We refer to this treatment as the "news article treatment" for ease of exposition, since this is likely how many respondents interpret it. The vignette is based on Di Tella and Rodrik (2020).

[8]See the online Appendix for additional results examining respondents' likelihood of making the same decision if they were the CEO of the company.

> This innovation could create new high-skilled jobs. These highly demanded high-skilled workers include those performing certain technical skills, required to deploy, operate and maintain the new digital technologies, specifically, AI, big data, and machine learning specialists. However, some low-skilled workers, specifically assembly and factory workers, who perform jobs with more repetitive tasks that can be easily automated, will lose their jobs to new machines and technologies. Furthermore, the remaining low-skilled workers will also see a cut in their yearly pay. The following tables show different possible scenarios.

Then, each individual saw four iterations of a conjoint table (see Table 1 for an example), each with randomly assigned combinations of the costs and benefits of automation. The online Appendix provides additional information on how the numbers presented in the conjoint tables were selected.

After viewing each table, respondents indicated whether they perceived the company's decision to introduce the new technology as fair. Finally, following the last table in the conjoint, respondents were asked how much they agreed or disagreed with each government policy listed in Section 4.1. This captured the aggregate effect of the specific information treatment across all four conjoint tables, since we did not ask the policy questions after each table.

5. Results

5.1 Perceived fairness of automation

The results of the between-group analyses of the perceived fairness of automation are shown in Figure 1.[9] Respondents in both the generic and specific trade-off conditions are more likely to view the company's automation decision as fair than respondents in the news condition, who were only exposed to the costs of automation. On a scale from 1 (strongly disagree) to 5 (strongly agree), the predicted value of the perceived fairness of automation is 3.23 [95 percent CI 3.18, 3.28] for the news condition, 3.68 [95 percent CI 3.63, 3.73] for the generic trade-off condition, and 3.37 [95 percent CI 3.35, 3.40] for the specific trade-off conjoint condition.

Contrary to what we expected, respondents in the generic information condition, who only saw vague information about the trade-offs of automation, are more likely to perceive the company's decision to automate as fair than individuals in the specific information condition, who saw precise estimates of the costs and benefits of automation. This suggests that in the absence of quantitative information, respondents may imagine automation to have a more positive impact than after having been exposed to the four specific scenarios in the conjoint experiment. To contextualize this, those in the conjoint condition, on average, were informed that the company's automation decision would result in: a fall in prices of about 21 percent, an increase of 67 high-skilled workers (from 200), and a decrease of 100 low-skilled workers (from 200), for a net loss of 33 workers; the average wage for high-skilled workers after the innovation was \$138,000 (from \$100, 000), an increase of 38 percent, whereas for low-skilled workers it was \$22,500 (from \$30,000), a decrease of 25 percent. This average scenario may entail greater costs or lower benefits than those inferred by respondents in the vague trade-off treatment, thereby leading to the perception of automation as being less fair.[10]

5.2 Support for policies addressing job loss from automation

Next, we turn to the effects of the treatments on support for policies addressing automation. Figure 2 shows that overall, the favored policy response across all groups is retraining workers. Conversely, the policies with the lowest support across all three treatment groups are those restricting skilled migration and imposing trade restrictions. This is largely consistent with

[9]See Table A1 in the online Appendix for the complete results.

[10]Alternative explanations for these findings are discussed in the online Appendix.

Table 1. Table shows the effects of the introduction of a productivity-improving innovation

Attributes	Before innovation	After innovation
Firm	Electronics *or* Aviation *or* Auto *or* Pharmaceutical	*Same as before innovation*
if Electronics: Price of [smartphone]	$600	$600 *or* $480 *or* $300
if Aviation: Price of [plane]	$100M	$100M *or* $80M *or* $50M
if Auto: Price of [car]	$25,000	$25,000 *or* $20, 000 *or* $12,500
if Pharmaceutical: Price of [vaccine]	$25	$25 *or* $20 *or* $12.50
Number of high-skilled workers	200	200 *or* 250 *or* 350
Wage of high-skilled workers	$100,000	$125,000 *or* $150,000
Number of low-skilled workers	200	150 *or* 50
Wage of low-skilled workers	$30,000	$20,000 *or* $25,000

Square brackets denote the pre-specified set of possible attribute values, contingent on the randomization of the firm attribute. For instance, if a respondent receives the "Auto" attribute for Firm, then only randomized pricing related to "car" ($25,000, $20,000, or $12,500) is shown. Words in *italics* do not appear in the conjoint experiment, and are included here for clarity.

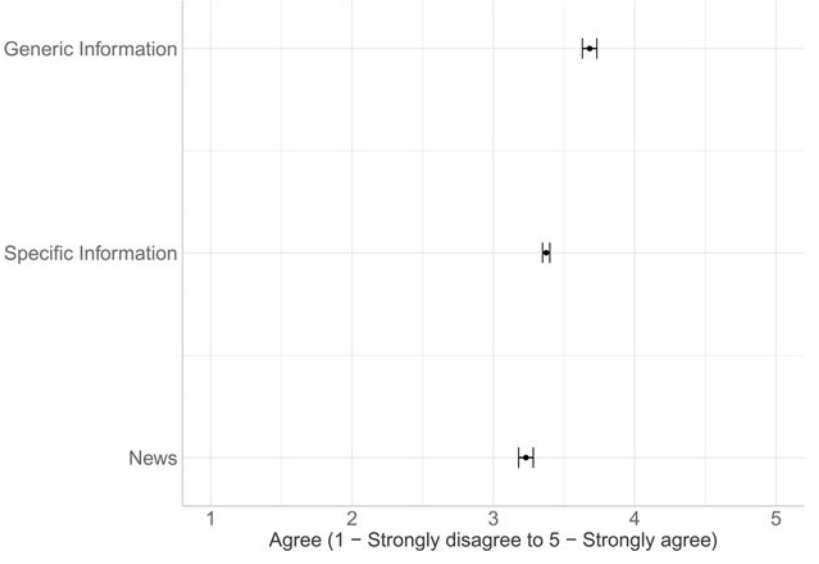

Figure 1. Perceived fairness of automation by treatment group; predicted values with 95 percent confidence intervals.

findings from Di Tella and Rodrik (2020) and Gallego *et al.* (2023), who find that respondents prefer unemployment compensation and training programs over import protection and do so by a significant margin.

In terms of the experimental treatment effects, we hypothesized that in response to automation-driven job losses, respondents in the specific information conjoint condition would be more likely to support social investment policies over compensatory policies and both of those policies over job protections, when compared to the other treatment groups. We expected individuals in the costs-only news condition to be more supportive of policies protecting jobs than those involving labor market interventions, compared to respondents in the other groups. Figure 2 shows that this is not the case, as there appear to be no differences in support for any of the policies across the treatment groups. Consistent with these descriptive results, between-group regression analyses of support for the eight individual policy responses show no effect of the generic and specific trade-off treatments relative to the costs-only news article condition (see Figures A8–A15 and Table A2 in the online Appendix for complete results). This suggests that regardless of the costs and benefits of automation, policy preferences are relatively sticky.

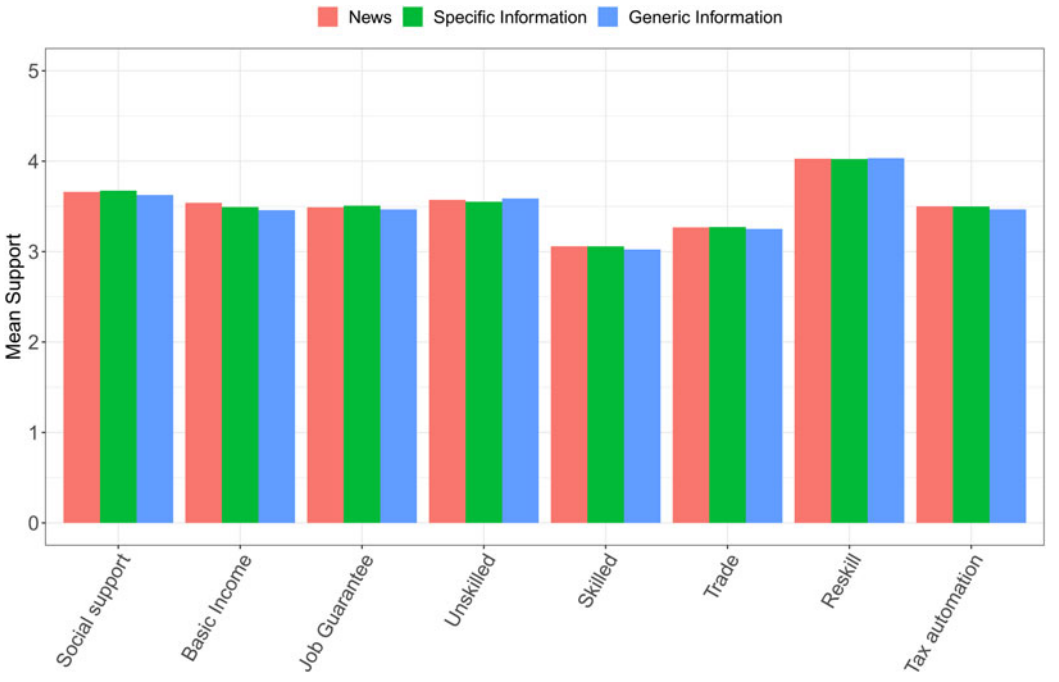

Figure 2. Mean levels of policy support by treatment group.

The absence of significant treatment effects on policy preferences fits with most existing experimental work on the subject. Several studies find null average treatment effects of automation primes on public policy preferences, although some report considerable heterogeneity across sociodemographic subgroups (Jeffrey, 2021; Ladreit, 2022; Wu, 2023; Zhang, 2022; Gallego *et al.*, 2023). While most of this research uses common treatments, whereby an automation shock affects one hypothetical firm (Di Tella and Rodrik, 2020; Ladreit, 2022; Wu, 2023; Zhang, 2022; Gallego *et al.*, 2023), and similar dependent variables, such as support for different policies in response to automation shocks, there is more variation in how the studies define the control groups: some compare automation shocks to an unspecified labor market shock, whereas others compare automation to particular shocks, such as offshoring. Our study, in contrast, compares a costs-only condition to treatments that emphasize both the costs and benefits—either vague or specific—of automation.

5.3 Conjoint analysis of perceived fairness of automation

Next we turn to the conjoint analysis that investigates which attributes causally increase or decrease the perceived fairness of the company's decision to automate. Figure 3 shows the results in terms of marginal means, which represent the mean outcome across all appearances of a particular conjoint feature level, averaging across all other features.[11] The most striking finding is that people are most sensitive to automation-driven changes in the prices of products.[12] Averaging across all other features, when prices decrease by 50 percent, the mean perceived fairness of automation is 3.52 on a 5-point scale [95 percent CI 3.48, 3.55], compared to 3.19 [95 percent CI 3.15, 3.22] when there is no price change.

[11]See Table A4 in the online Appendix for complete results.
[12]See Figure A16 in the online Appendix for average marginal component effects.

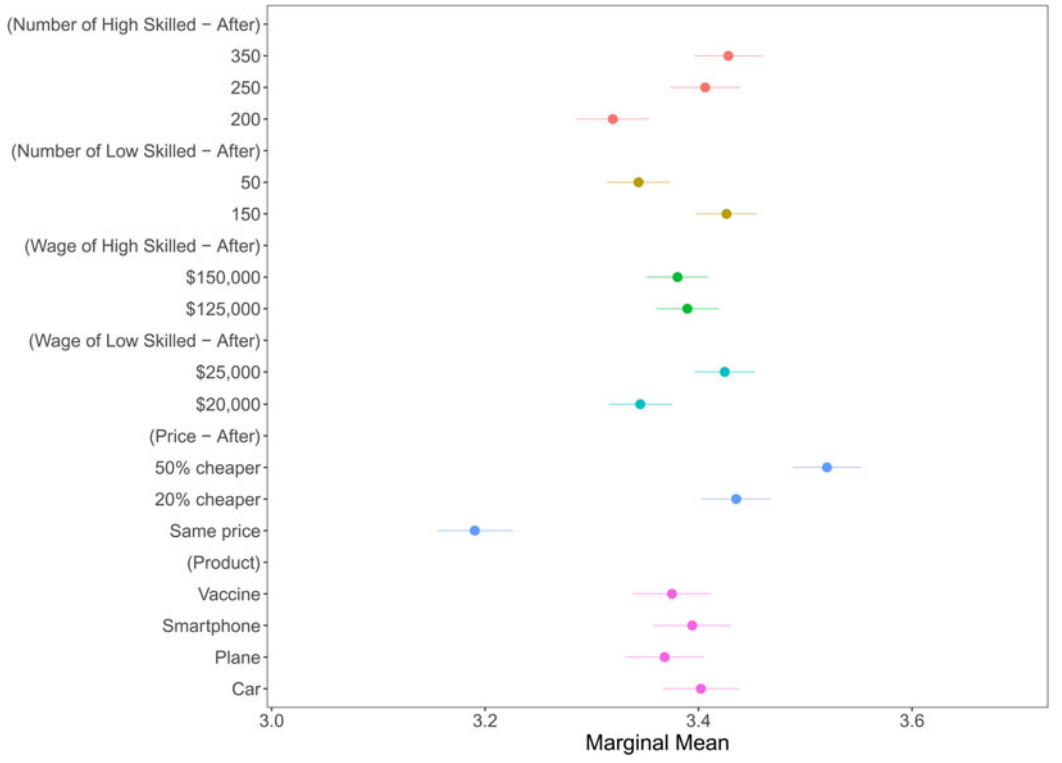

Figure 3. Marginal means of conjoint experiment features on perceptions of fairness.

Similarly, when fewer (more) workers are lost (gained), respondents are more likely to think that the firm's decision to automate is fair, compared to when more (fewer) workers are lost (gained). They are also sensitive to varying declines in the wages of workers affected by automation, but not to wage increases, which suggests that perceptions of gains and losses are not symmetric. This is consistent with prospect theory, which argues that people are more sensitive to economic losses than economic gains (Kahnemann and Tversky, 1979). Finally, respondents do not appear to be sensitive to the type of firm or product in question, whether it is an auto company producing cars, an aviation firm producing planes, an electronics company making smartphones, or a pharmaceutical company producing vaccines.

Due to power limitations, we cannot estimate the marginal means for the complete best-case scenario compared to the worst-case scenario, whereby we set each profile feature to its most optimistic or most pessimistic value, but we can do so for up to three features. When we set the price decrease to 50 percent, the number of high-skilled jobs post-innovation to 350 (the largest increase) and low-skilled jobs post-innovation to 150 (the lowest decrease), averaging across all other features, the mean perceived fairness of automation is 3.60 [95 percent CI 3.54, 3.67]. Conversely, when we set the price decrease to 0 percent, the number of high-skilled jobs post-innovation to 200 (no change) and low-skilled jobs post-innovation to 50 (the largest decrease), averaging across all other features, the mean perceived fairness of automation is 3.09 [95 percent CI 3.02, 3.16] (see Table A3 in the online Appendix for complete results). The point estimate of the best-case scenario is roughly equivalent to that of the generic information group, suggesting again that the net benefits inferred by respondents in the generic trade-off condition are higher than those of the average conjoint scenario. Finally, we want to note the substantive significance of this finding. A shift from a worst-case scenario to a best-case scenario moves people from neutral to weak support for automation, which could be politically consequential (as it could shift

voters from not caring to caring about the issue—even if only slightly) and is a fairly large attitudinal change for a political behavior study. It is also worth keeping in mind that the observed effect is generated by a one-time experimental manipulation in a setting without any additional information. Automation is not currently a salient political issue (Gallego and Kurer, 2022) but presumably, were politicians to seize on it and amplify it using persistent, well-crafted messaging about the fairness (or lack thereof) of automation-driven labor market changes, the resulting public opinion effects could be greater than those observed in our study.

6. Conclusion

Although projections of the future economic impact of technological change vary, automation and AI are likely to drastically transform the nature of work. Given that these labor market changes are bound to have far-ranging distributive consequences, scholars are increasingly interested in the implications of automation and AI for politics. Workers negatively impacted by technological change may, for instance, become disaffected and support more radical forms of politics if governments fail to adequately compensate them. But the negative economic consequences of technological change are only one side of the story. Automation-driven labor market transformations will also create new employment opportunities, perhaps partly offsetting its destabilizing effects. Despite the looming consequences of technological change, most political parties have not claimed ownership of automation or formulated specific policies in response to it; instead they have found it more politically advantageous to focus their campaigns on more traditional economic shocks, such as trade and immigration, which offer more convenient scapegoats and seemingly simpler solutions. At the same time, the scholarly literature has primarily focused on automation's negative consequences (especially job loss) and has not explored how people respond to information about its potential benefits, such as job gains and price decreases. In this paper, we have sought to address these gaps in existing research. Do people perceive a firm's decision to automate to be fairer if information on both its costs and benefits is disclosed? Does providing information about the trade-offs of automation affect support for policies that address automation's negative consequences?

Overall, our findings suggest that relative to messages that only convey the costs of automation, the perceived fairness of automation increases significantly when generic, and to a lesser extent specific, information about the trade-offs of automation is provided. People are particularly sensitive to price changes: as the prices of consumer products decrease due to automation, the perceptions of its fairness increase. Finally, the policy response to automation that garners the greatest support among respondents is the retraining of individuals negatively affected by technological change.

Our results suggest that whether technological change generates political backlash, as has been anticipated by many commentators, depends on the degree to which its costs are offset, or even exceeded, by its benefits. Thus, how political parties choose to frame automation—for instance, whether they emphasize its price-reducing potential or its positive implications for labor markets—will affect voters' perceptions of the issue and the degree to which they see it as a political problem or an economic opportunity.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2024.1. To obtain replication material for this article, https://doi.org/10.7910/DVN/AJGHMH

Acknowledgements. We are grateful for feedback we received after presenting earlier versions of this work at the University of Washington, the California Institute of Technology, the Policy, Elections, and Representation Lab at the University of Toronto, the Schwartz Reisman Institute for Technology and Society, and the University of British Columbia, Okanagan. The authors would like to thank Chris Adolph, Michael Alvarez, Victor Menaldo, and Lee Slinger for their helpful comments and valuable suggestions on previous versions of the article. Finally, the authors would also like to thank the editor and two anonymous reviewers, whose comments greatly improved the paper. Errors are our own.

Funding statement. This study has been supported by the Schwartz Reisman Institute for Technology and Society.

Competing interests. None.

References

Abbott R and Bogenschneider B (2018) Should robots pay taxes? Tax policy in the age of automation. *Harvard Law and Policy Review* **12**, 145–175.

Aghion P, Antonin C, Bunel S and Jaravel X (2020) What are the labor and product market effects of automation? New evidence from France. Working paper.

Aghion P, Antonin C, Bunel S and Jaravel X (2021) The direct and indirect effects of automation on employment: a survey of the recent literature. Working paper.

Alesina A and Angeletos G-M (2005) Fairness and redistribution. *American Economic Review* **95**, 960–980.

Alesina A and Giuliano P. (2011) Preferences for redistribution. In Benhabib J, Bisin A and Jackson MO (eds). *Handbook of Social Economics, 1A of Handbook of Economics*. Burlington: Elsevier Science, pp. 93–131.

Anelli M, Colantone I and Stanig P (2019) We were the robots: automation and voting behavior in Western Europe. IZA Discussion Paper Series. No. 12485.

Autor DH and Dorn D (2013) The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review* **103**, 1553–1597.

Autor DH, Levy F and Murnane RJ (2003) The skill content of recent technological change: an empirical exploration. *The Quarterly Journal of Economics* **118**, 1279–1333.

Autor DH, Mindell DA and Reynolds EB (2020) *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. Cambridge, MA: The MIT Press.

Baker A (2003) Why is trade reform so popular in Latin America? A consumption- based theory of trade policy preferences. *World Politics* **55**, 423–455.

Berinski AJ, Huber GA and Lenz GS (2012) Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis* **20**, 351–368.

Bessen J (2019) Automation and jobs: when technology boosts employment. *Economic Policy* **34**, 589–626.

Borwein S, Bonikowski B, Loewen PJ, Lee-Whiting B and Magistro B (2024a) Perceived automation threat, populism, and vote choice: evidence from 15 European democracies. *West European Politics*. https://doi.org/10.1080/01402382.2023.2297601.

Borwein S, Bonikowski B, Loewen PJ, Magistro B and Lee-Whiting B (2024b) Who can assert ownership over automation? Workplace technological change, populist and ethno-nationalist rhetoric, and candidate support. *Political Behavior*. https://doi.org/10.1007/s11109-024-09914-0.

Busemeyer MR and Tober T (2023) Dealing with technological change: social policy preferences and institutional context. *Comparative Political Studies* **56**, 968–999.

Chatruc MR, Stein E and Vlaicu R (2021) How issue framing shapes trade attitudes: evidence from a multi-country survey experiment. *Journal of International Economics* **129**, 103428.

Ciccone A, Rogeberg O and Braaten R (2020) Fairness preferences in a bilateral trade experiment. *Games* **11**, 1–8.

Coppock A and McClellan OA (2019) Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics* **6**, 1–14.

Dauth W, Findeisen S, Suedekum J and Woessner N (2021) The adjustment of labor markets to robots. *Journal of the European Economic Association* **19**, 3104–3153.

Dermont C and Weisstanner D (2020) Automation and the future of the welfare state: basic income as a response to technological change?. *Political Research Exchange* **2**, 1–11.

Di Tella R and Rodrik D (2020) Labour market shocks and the demand for trade protection: evidence from online surveys. *The Economic Journal* **130**, 1008–1030.

Frey CB and Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation?. *Technological Forecasting and Social Change* **114**, 254–280.

Frey CB, Berger T and Chen C (2018) Political machinery: did robots swing the 2016 US presidential election?. *Oxford Review of Economic Policy* **34**, 418–442.

Gallego A and Kurer T (2022) Automation, digitalization, and artificial intelligence in the workplace: implications for political behavior. *Annual Review of Political Science* **25**, 463–484.

Gallego A, Kuo A, Manzano D and Fernández-Albertos J (2021) Technological risk and policy preferences. *Comparative Political Studies* **55**, 60–92.

Gallego A, Kurer T and Schöll NB (2022) Neither left-behind nor superstar: ordinary winners of digitalization at the ballot box. *Journal of Politics* **84**, 418–436.

Gallego A, Kuo A and Manzano D (2023) Automation versus openness: support for policies to address job threats. *Journal of Public Policy*, 1–23. doi: 10.1017/S0143814X23000260.

Goos M, Manning A and Salomons A (2014) Explaining job polarization: routine-biased technological change and offshoring. *American Economic Review* **104**, 2509–2526.

Han S and Kwon HY (2020) Employment insecurity and social policy: preferences for investment vis-à-vis consumption. *Policy and Society* **39**, 247–265.

Hemerijck A (2018) Social investment as a policy paradigm. *Journal of European Public Policy* **25**, 810–827.

Huff C and Tingley D (2015) "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics* **2**, 1–12.
Humlum A (2019) Robot adoption and labor market dynamics. Working paper.
Im ZJ (2021) Automation risk and support for welfare policies: how does the threat of unemployment affect demanding active labour market policy support?. *Journal of International and Comparative Social Policy* **37**, 76–91.
Im ZJ, Mayer N, Palier B and Rovny J (2019) The "losers of automation": a reservoir of votes for the radical right?. *Research & Politics* **6**, 1–7.
Jeffrey K (2021) Automation and the future of work: how rhetoric shapes the response in policy preferences. *Journal of Economic Behavior and Organization* **192**, 417–433.
Jeffrey K and Matakos K (2021) Economic anxiety, anti-elite sentiment, and demand for redistribution. QPE working paper.
Kahneman D, Knetsch JL and Thaler R (1986) Fairness as a constraint on profit seeking: entitlements in the market. *American Economic Review* **76**, 728–741.
Kahnemann D and Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–292.
Kurer T (2020) The declining middle: occupational change, social status, and the populist right. *Comparative Political Studies* **53**, 1798–835.
Kurer T and Gallego A (2019) Distributional consequences of technological change: worker-level evidence. *Research & Politics* **6**, 1–9.
Kurer T and Häusermann S (2011) Automation risk, social policy preferences, and political participation. In Busemeyer MR, Kemmerling A, Marx P and Van Kersbergen K (eds), *Digitalization and the Welfare State*. Oxford University Press.
Ladreit C (2022) Automation and public policy preferences. Working Paper.
Magistro B (2022) Party cues or policy information? The differential influence of financial and economic literacy on economic policy preferences. *Journal of Public Policy* **42**, 465–488.
Margalit Y (2013) Explaining social policy preferences: evidence from the great recession. *American Political Science Review* **107**, 80–103.
Martinelli L (2020) A basic income trilemma: affordability, adequacy, and the advantages of radically simplified welfare. *Journal of Social Policy* **49**, 461–482.
Marx P (2014) The effect of job insecurity and employability on preferences for redistribution in Western Europe. *Journal of European Social Policy* **24**, 351–366.
Mullinix KJ, Leeper TJ, Druckman JN and Freese J (2015) The generalizability of survey experiments. *Journal of Experimental Political Science* **2**, 109–113.
Mutz D (2021) (Mis)Attributig the causes of American job loss: the consequences of getting it wrong. *Public Opinion Quarterly* **85**, 101–122.
Naoi M and Kume I (2015) Workers or consumers? A survey experiment on the duality of citizens' interests in the politics of trade. *Comparative Political Studies* **48**, 1293–1317.
Neimanns E, Busemeyer MR and Garritzmann J (2018) How popular are social investment policies really? Evidence from a survey experiment in eight Western European countries. *European Sociological Review* **34**, 238–253.
Rodrik D (2018) Populism and the economics of globalization. *Journal of International Business Policy* **1**, 12–33.
Rubin JD (2012) Fairness in business: does it matter, and what does it mean?. *Business Horizons* **55**, 11–15.
Sacchi S, Guarascio D and Vannutelli S (2020) Risk of technological unemployment and support for redistributive policies. In Careja R, Emmenegger P and Giger N (eds), *The European Social Model under Pressure*. Wiesbaden: Springer VS, pp. 277–295.
Schöll N and Kurer T (2024) How technological change affects regional voting patterns. *Political Science Research and Methods* **12**, 94–112.
Starmans C, Sheskin M and Bloom P (2017) Why people prefer unequal societies. *Nature Human Behaviour* **1**, 1–7.
Thewissen S and Rueda D (2019) Automation and the welfare state: technological change as a determinant of redistribution preferences. *Comparative Political Studies* **52**, 171–208.
Werfel S, Witko C and Heinrich T (2022) Public support for assistance for workers displaced by technology. *Research & Politics* **9**, 1–7.
Wu N (2021) Misattributed blame? Attitudes toward globalization in the age of automation. *Political Science Research and Methods* **10**, 470–487.
Wu N (2023) Restrict foreigners, not robots': partisan responses to automation threat. *Economics & Politics* **35**, 505–528.
Zhang B (2022) No rage against the machines: threat of automation does not change policy preferences. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES'22)*, August 1–3, 2022, Oxford, United Kingdom. New York, NY, USA: ACM, 11 pages. https://doi.org/10.1145/3514094.3534179.

Cite this article: Magistro B, Loewen P, Bonikowski B, Borwein S, Lee-Whiting B (2024). Attitudes toward automation and the demand for policies addressing job loss: the effects of information about trade-offs. *Political Science Research and Methods* **12**, 783–798. https://doi.org/10.1017/psrm.2024.1