University of Duisburg-Essen

- Institute for Political Science –

Master's Thesis:

The Social Identities of Biden and Trump Sympathizers – A Multilevel-Analysis of the Candidate Preference in the U.S. Presidential Election 2020

First Examiner: Paul Vierus, MSc

Second Examiner: Prof. Dr. Achim Goerres

Word Count: 20.873

Characters: 138.618 (incl. blanks)

Manuel Diaz Garcia
Blumenthalstr. 56a, 47058 Duisburg
E-mail: manuel.diaz-garcia@stud.uni-due.de
Student ID: 3033979
Duisburg, 10.10.2022

Abstract

The U.S. Presidential Election 2020 was a contest between two candidates referring to different social identities during their campaigns. While Donald Trump raised the salience of white, nationalist, and Christian identities, Joe Biden emphasized more ethnically diverse, less nationalist, and secular identities. Considering the state context in an election that builds upon a winner-take-all principle at the state level, this thesis uses a novel approach to explain candidate preference in the 2020 election. Data from the American National Election Study 2020 is used in combination with contextual data for a multilevel analysis of preference between Donald Trump and Joe Biden. National and religious identity work in the expected ways. However, a stronger white identity was associated with a higher likelihood of preferring Joe Biden, contradicting theoretical expectations and previous research. The identity effects hold even when sociotropic economic considerations are included. Additionally, ethnic diversity and partisan lean at the state level affected candidate preference in the 2020 election.

Acknowledgments: I want to thank Paul Vierus for his excellent supervision of this master's thesis as well as the colloquium led by Achim Goerres and Conrad Ziller for fruitful feedback on important parts of the theory. Special thanks to Dylan Wojtkowski for grammar, style, and spelling corrections.

Contents

1. Introduction

This thesis focuses on candidate preference in the U.S. Presidential Election 2020. A multilevel approach will incorporate Social Identity Theory at the individual level and various factors at the contextual level influencing the candidate preference of people in the United States.

In 2000 George W. Bush won the U.S. Presidential Election by a margin of 537 votes after the U.S. Supreme Court eventually had to stop a recount in the state of Florida (Elving 2018). In 2016 Hillary Clinton lost the election by around 130.000 votes in Florida and Wisconsin, making Donald Trump the 45th President of the United States (New York Times 2017). And in 2020, Donald Trump lost his reelection bid by around 40.000 votes in Arizona, Georgia, and Wisconsin (Lindsay 2020). In all these elections, only a few votes in the right states determined the fate of the whole nation. Due to the nature of the U.S Presidential Election – the winner-take-all principle at the state level and the Electoral College deciding on who becomes President – a few thousand votes in a few states can determine the outcome of an election where the voting-age population is over 200 million people in 50 different states and the District of Columbia.

The 2020 U.S Presidential Election was an exceptional situation that is worth looking at for two reasons. First is the historical scope of the election. Historian Bruce J. Schulman described it as a "momentous episode in the story of American democracy" (Schulman 2020). The largest total voter turnout in history was expected, COVID-19 caused an unprecedented voting environment during a global pandemic, and innovative ways of voting like absentee and mail-in ballots gained in usage. Second, the 2020 election dealt with the most polarized electorate in history. The antagonism between the Democratic Party and the Republican Party and their members – or affective polarization – has shown to be one of the most critical factors in the 2020 election (Garzia and Ferreira da Silva 2022: 303). Liliana Mason linked this polarization between the parties to social identities like ethnic, ideological, or religious identities, which sorted themselves into one party and developed mega-partisan identities (Mason 2016: 10).

Furthermore, in a highly personalized election like the U.S. Presidential Election, one of the most powerful predictors of vote choice is which candidate appeals to more voters (Campbell et al. 1960: 39) – or in other words, which candidate do people prefer? Previous research has focused on various aspects when studying candidate preference. From authoritarianism (Luttig 2021) and issue polarization (Ocampo, Garcia-Rios, and Gutierrez 2021) to religious practices (Bradberry 2016) or political and institutional environments (André and Depauw 2017).

Nevertheless, the essential characteristics of the U.S. Presidential Elections mentioned above are not considered comprehensively and systematically. The importance of the state level in U.S. Presidential Elections, social identities in contemporary American politics, and candidate preference in election outcomes is undeniable. However, Political Science has still failed to study candidate preference from a social identity perspective and with a multilevel approach focusing on the state level. The thesis will close this research gap by answering the following two research questions regarding the U.S. Presidential Election 2020:

> **RQ1:** *To what extent do social identities influence the candidate preference between Joe Biden and Donald Trump in the U.S. Presidential Election 2020?*
>
> **RQ2:** *How relevant is the state-level in explaining candidate preference in the U.S. Presidential Election 2020, and which characteristics are most important?*

These research questions will be answered in four steps. First, a literature review will point out the gap in existing research, which is aimed to be closed. Second, a theoretical model will be developed rooted in Social Identity Theory, the convergence of identities in recent years, and how the candidates manage to activate specific social identities. Third, the different datasets and methods will be explained. Multilevel logistic regression with individual-level data from the American National Election Study 2020 and contextual-level data from other sources are used to get insights into the effects of social identities and contextual-level predictors on candidate preference. Fourth, the multilevel analysis results will be reported, discussed, and criticized in the context of existing research.

The analysis provides three main findings. First, a stronger white identity is associated with an increased likelihood of preferring Joe Biden. In comparison, a stronger national identity and a Christian identity are associated with an increased chance of preferring Donald Trump. Second, the state level proved important in explaining candidate preference in the U.S Presidential Election 2020. The variance in candidate preference at the state level was almost entirely resolved when contextual level predictors were included. Third, Lower ethnic diversity and living in a Democratic-leaning state were associated with an increased likelihood of preferring Joe Biden. In contrast, higher ethnic diversity and living in a Republican-leaning state were associated with an increased likelihood of preferring Donald Trump.

2. Literature Review

2.1. Social Identities in Electoral Research

The field of electoral research is defined by the dependent variables it tries to explain (van der Eijk 2002: 189). Political attitudes like opinions on different policy issues, political interest, and candidate preference, as well as political behavior like vote choice or turnout, are the most central topics of electoral research. The most fundamental work in electoral research is Campbell and colleagues' 1960 "The American Voter", which led the whole field in a new direction. For the first time, individual-level characteristics of voters and psychological processes played a major role in explaining electoral behavior in the United States (Campbell et al. 1960: 17). What became famous as the "Michigan Model" focused on party identification, issue orientation, and candidate preference as predictors for vote choice (Campbell et al. 1960: 66). Since then, multiple other psychological factors have joined the attempt to explain political attitudes and political behavior.

One theoretical framework that enjoyed special attention is Tajfel and Turner's Social Identity Theory from 1979. Numerous variables regarding electoral research have been explained by multiple social identities like ethnic, national, or partisan identities. In the U.S. context, ethnic identity has been used to describe political behavior among Latinos. Jackson finds an effect of priming the ethnic identity of Latinos while also having a co-ethnic candidate in the 2006 California Gubernatorial Election on vote preference (Jackson 2011: 712). According to Social Identity Theory, group identities like Latino identity get more salient when activated through priming and impact political attitudes and behavior (Jackson 2011: 694). Perez finds similar results in a survey experiment. First, Latinos are more likely to engage politically when they feel statements of political elites threaten their identity. Second, strong identifiers are even more motivated to defend their in-group (Perez 2015: 174). These findings follow Social Identity Theory since groups try to maintain a more positive picture of themselves when challenged (Perez 2015: 160). In this case, Latinos try to keep a positive feeling toward the ingroup with political engagement after disrespectful statements by political elites. Adida and colleagues studied the effects of ethnic identity on the ability to recall presented information on political candidates correctly in a more ethnically diverse context. In an experimental setting in Benin, they found that positive information about a candidate only had positive effects on coethnics while negative information negatively affected non-coethnics (Adida et al. 2017: 469). These results are explained by motivated reasoning, which suggests a bias in information processing based on the candidate's ethnic group (Adida et al. 2017: 441).

National Identity and its effect on voting behavior has been a major research topic in Europe over the last decade. Amid the rise of right-wing parties in European countries, the extent to which people feel attached to their nation has become more important in European politics (Lubbers and Coenders 2017: 115). Authors argue that people turn to right-wing parties to preserve their national identity, which they perceive as threatened by different factors, e.g., more migrants coming to Europe or the EU getting more competencies relative to the nation-states (Lubbers and Coenders 2017: 100). This effect also appeared during the Brexit-Referendum in the UK in 2016. Perceived threats to the national identity increased the likelihood of voting "Leave" (Swami et al. 2018: 170).

Partisan identity, the sense of feeling psychologically attached to a particular party, is perhaps the most studied identity when it comes to political attitudes and behavior. The effects of instrumental (based on policy preferences) and expressive partisanship (based on social identities acquired early in life) have shown to be important in the U.S. as well as in Europe (Huddy, Bankert, and Davies 2018: 174). In the U.S. Congressional Elections 2010, a threatened partisan identity led to higher turnout (Miller and Johnston Conover 2015: 236). Since elections can be perceived as group competition between parties, they motivate partisans to turn out if a potential loss to the rival party threatens their identity (Miller and Johnston Conover 2015: 228). For the German Bundestagswahl 2009, Stefanie Reher finds that partisanship motivates voters to turn out regardless of issue priorities due to higher self-esteem and satisfaction from electoral success (Reher 2014: 8). Moreover, partisan identity also seems to shape the news perception of voters during election campaigns. For the U.S. Presidential Election campaign in 2004, Hoffner and Rekhoff found differences in news perceptions between Democrats and Republicans. Threatened partisan identity by negative news coverage was stronger among Republicans. However, Republicans thought that other people were more influenced by news coverage than themselves (Hoffner and Rehkoff 2011: 749-50).

Some other identities are used to explain different kinds of political attitudes and political behavior. Devine showed in an online experiment in the U.S. that self-placement effects on vote choice only exist for people with at least a moderate attachment to their ideology of being conservative or liberal (Devine 2015: 530). In Denmark, educational identity, instead of just group membership, seems to matter for vote choice. High-education identifiers prefer social liberals while low-education identifiers prefer conservatives (Stubager 2013: 388). In addition, Andersen and Heath look at the different political contexts of the U.S., UK, and Canada and hypothesize different social identities shaping political attitudes in each context and find differing effects of age- and religion-based identities (Andersen and Heath 2003: 303-04).

Social identities in the context of electoral research have proven to be successful indicators for explaining political attitudes and political behavior. However, most studies only focus on one political or institutional context. Aside from rare exceptions, a multilevel perspective is missing from research on the effects of social identities on political attitudes and behavior. In addition, the potential effects social sorting might have on political attitudes and behavior are overlooked. The importance of converging social identities like ethnic, national, and religious identity to a mega-partisan identity and the consequences for American elections will be pointed out later. Since the thesis focuses on the effects of certain social identities on candidate preference in the U.S. Presidential Election 2020 the following section will review the previous research on candidate preference.

2.2. Previous Research on Candidate Preference

Previous research on candidate preference during elections has focused on candidate preference as an independent and dependent variable. In the context of how social identities influenced candidate preference in the U.S. Presidential Election 2020, only the literature that uses candidate preference as a dependent variable is relevant. Multiple factors that might explain candidate preference have been studied: From social identities over other individual-level characteristics to contextual-level determinants.

Studies focusing on social identities and their influence on candidate preference examine the effects in different contexts. King looks at racial and gender identities in the Democratic Primaries 2020. He shows that self-categorization and comparisons with a candidate's group membership are important for voters to prefer a candidate (King 2022: 10). Additionally, Kam shows that attitudes toward a social group shape the willingness to support a minority candidate. Although, this effect fades if the information on a candidate's party affiliation is presented (Kam 2007: 362). Partisanship moderates the effects of group attitudes (Kam 2007: 348). In the 2001 Los Angeles Mayoral Election, candidate preference for a Latino candidate was higher for strong identifying Latinos (Stokes-Brown 2006: 642). However, if white people report preferring a candidate of their social group this automatically comes with condemnation of potential prejudice they might hold against candidates of other groups (Moss-Racusin, Phelan, and Rudman 2010: 553). König and Waldvogel (2022) looked at party identity and how this shaped candidate preference during TV debates during the 2017 German Federal Election campaign. They argued that party identity acts as a filter for how issue statements of candidates are understood and found that only 10% of participants changed their candidate preference during the TV debates (König and Waldvogel 2022: 314).

In addition to social identities, various individual-level determinants found their way into explaining candidate preference in different types of elections. Authoritarianism came with a higher candidate preference for the Republican candidate in the 2016 U.S. Presidential election (Luttig 2021: 786). In the 2020 U.S. Presidential Election, issue-polarization on COVID-19 and the Economy shaped the different preferences for Donald J. Trump and Joe Biden among Latinos (Ocampo, Garcia-Rios, and Gutierrez 2021: 16). During the Republican Primaries in 2008, the candidate with the most religious appeals performed best among voters with the most religious practices (Bradberry 2016: 21). And a study by Rhodebeck analyzing the U.S. Presidential Elections from 1972 to 1988 shows Republicans gain more from elderly conservatives than Democrats do from elderly liberals (Rhodebeck 1993: 361).

While individual-level characteristics are crucial in studying candidate preference, contextual-level determinants are still mostly overlooked. McCann (2009) looks at the perceived threat to the own way of life at the state level in the Congressional Elections 1946-1992. In this time-series study, higher societal threat in a state was associated with supporting the Republican candidate (McCann 2009: 350). In Europe, Andre and Depauw (2017) hypothesize the effects of the political and institutional environment on the propensity to vote for different parties. Although the effects are more common among politically more sophisticated voters (André and Depauw 2017: 613), they still emphasize the importance of institutional context for candidate preference (André and Depauw 2017: 599-600).

The review of previous research has shown that two pieces are missing for a comprehensive picture of candidate preference in the U.S. as part of the electoral research area. First, the lack of a systematic analysis of social identities' effects on candidate preference that play a crucial role in the process of convergent identities. Second, contextual factors are mostly missing when studying candidate preference. Most explanatory variables only focus on the individual level. Therefore, this thesis uses Social Identity Theory in combination with the idea of converging identities to a mega-partisan identity to explain candidate preference in the U.S. Presidential Election 2020 in a multilevel framework to fill this research gap.

3. Theory

3.1. Social Identity Theory and Mega-Partisan Identities

The main theoretical arguments for the individual-level determinants of candidate preference in the U.S. Presidential Election in 2020 are built upon Tajfel & Turner's Social Identity Theory. The theory was developed in two essential steps. The first step was the joint work of Tajfel and Turner (1979), who developed the Social Identity Theory of Intergroup Behavior. The second step was Turner's creation of the so-called Social Identity Theory of the Group (or Social Categorization Theory), a more general approach than the theory developed in 1979. But to fully understand the concept of social identities, derive hypotheses from it, and detect possible consequences, one must look at both works as closely related.

In the sense of Social Identity Theory "Social Identity [is] understood as the part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership" (Tajfel 1981: 255). This comprehensive definition touches on three critical aspects of social identities. First is the self-concept as a cognitive psychological process referred to as self. Every individual possesses self-concepts that are also context-dependent (Turner 1987: 44). The context-dependency and saliency of social identities will become important later. Second are the value and emotional significance of a social group. This value and significance are relational to other groups. Social comparisons with other groups lead to positive or negative perceptions of the own group and, therefore, the self-concept (Tajfel and Turner 1979: 40). For the value and significance of a social group to develop within individuals, a sense of belonging to that group is necessary. Only being an objective group member is not sufficient (Huddy 2013: 738). Third is the membership in a social group that is only achieved if an individual perceives himself as part of that group. Oneself must share the same social category and emotional involvement with others (Tajfel and Turner 1979: 40).

Turner's Social Categorization Theory (SCT) explains the social-psychological processes underlying group behavior. The central concept for SCT is the phenomenon described as *Depersonalization*. This means a change from an individual-level identity to a group-level identity where self-stereotyping leads individuals to perceive themselves more as part of a social group than as unique human beings (Turner 1987: 50). Moreover, the process of Depersonalization has multiple consequences on the perception of oneself and others. Self-categories tend to be evaluated positively, oneself and others are assessed based on group comparisons, and in-group members are evaluated more positively than out-group members

(Turner 1987: 57). Therefore, positive social identity is a result of the perceived positive distinctiveness of one's own group from other groups (Turner 1987: 62).

Individuals striving for this positive distinctiveness of their social group is the key to Tajfel and Turner's Social Identity Theory of Intergroup Behavior. The positive distinctiveness is created through positive connotations of a group. Positive connotations, in turn, develop from multiple comparisons with out-groups in which the in-group is favored in some way (Tajfel and Turner 1979: 40). In the case of a negative social identity, group members have two options: (1) change the interpretation of their group and try to make more positive social comparisons, or (2) accept the inferiority and try to take action for social change (Tajfel 1981: 256). However, all these things only play a role if there are relevant outgroups for comparisons and when the social identities become salient.

Since social identities heavily rely on social comparisons, identities need to be salient to affect social attitudes and behavior because only salient identities can be compared (Tajfel and Turner 1979: 41). This leads to group-based action only being important in social situations that activate the psychological attachment to a group. Some identities are more salient in general, while others only get activated under certain conditions (Tajfel 1981: 255). According to Turner, saliency is achieved when both in-group and out-group members are part of a group comparison in a social situation. This leads to individuals focusing on in-group similarities and out-group differences; the more comparisons are made, the stronger social identities get (Turner 1987: 49). Since different reference groups become important in different social situations, the effects of social identities are very much context-based because different identities are salient depending on the circumstances (Turner 1987: 46). Ethnic identity might play a crucial role in social attitudes and behavior in a highly diverse country which allows for multiple group comparisons, while it may be unimportant in a more homogeneous country.

While Tajfel and Turner show how social identities affect social attitudes and behavior, a multitude of researchers developed theoretical frameworks for social identities explaining political attitudes and behavior (see Pauls, Leibon, and Rockmore 2015; Greene 2004; Ben-Bassat and Dahan 2012). The question arising at this point is how social identities become relevant for political actions. Leonie Huddy argues that social identities acquire political relevance if group members develop a set of shared political beliefs, which results in support for the same political parties, candidates, or policies. If these actions can be attributed to the social identity of individuals, she calls this group-based political cohesion (Huddy 2013: 739). This politicization of social identities leads to various consequences for the political world.

Perceived threats to a group's status (Huddy 2013), different intensities of identities (Green and Cowden 1992), partisan biases (Bartels 2002; Malka and Lelkes 2010), saliency in different social situations (Transue 2007), and convergent identities (Mason 2012) can all impact how individuals behave politically.

Especially the effects of convergent identities, pointed out by Liliana Mason, have become important in the U.S. in recent years. Since the 1960s, the concept of cross-pressures was how scholars described the situation in the American electorate. Partisan identifiers were mostly part of social groups that conflicted with their party identity (Campbell et al. 1960; Powell 1976). These cross-cutting identities led to more contact with members of the other partisan group, which in turn created social pressures leading to bipartisan understanding (Cassese 2020: 170) – the so-called cross-pressures. However, the positive effects of cross-pressures like social understanding and members of different partisan groups getting along faded in recent years due to increased social sorting (Mason 2016: 5). Social sorting describes the development of members of one social group also becoming members of other social groups (Mason 2016: 4). Mason based her work mainly on the sorting of partisan and ideological identities which resulted in a correct combination of these two identities (Mason 2012: 143). Democratic and liberal identities aligned, while Republican and conservative identities did the same. The country moved from a political landscape of cross-cutting identities to aligned identities (Mason 2012: 144). Partisan strength, ingroup biases, activism, and anger increased in the U.S. society, and tolerance of the other party became less likely (Mason 2012: 143-44) because "an aligned identity is a stronger identity" (Mason 2015: 131). As a result, people with highly sorted identities tend to interact mostly within their social groups and lack inter-group contact (Mason 2016: 5). Ideology is not the only identity that started to align with partisan identities. Research has shown that identities like religion (Layman 1997: 306) or race (Giles and Hertz 1994: 322) also align with partisan identity. Since cross-pressures that could have moderated social conflict between partisans were mainly left behind, the electorate grew increasingly emotionally engaged in politics, and a new identity developed, which Mason calls the mega-partisan identity (Mason 2016: 3). The result is a more affectively polarized electorate.

Mega-partisan identities arise when all the different social identities, which are not genuinely political, align with the partisan identity. As mentioned above, communication gets more difficult if cross-pressures get lost between partisan groups. More social isolation, emotional reactions to political events, and stronger emotional reactions to policy issues are the result of mega-partisan identities (Mason 2016: 18). The mega-partisan identity is the convergence of the usually studied partisan identity with other identities like ideology, race, religion, or nationality. For example, according to Mason, the identities "black", "liberal", and "secular"

sorted themselves into the Democratic Party, while "white", "conservative", and "evangelical" are linked to the Republican Party (Mason 2016: 10). National identity, however, seems to play a different role here. It has long been thought of as unideological (Huddy and Khatib 2007: 65). It has shown to be effective in reducing affective polarization by referring to a common group identity of partisans (Levendusky 2017: 1).

Aside from obviously political identities like partisan and ideological identities, the most important identities that play a role in convergent identities are ethnic, national, and religious identities. And they are not genuinely political identities. However, they became part of the newly developed mega-partisan identities because supporters of the two major parties in the U.S. started using social identities to differentiate between the two parties (Ahler 2018: 19). These parties no longer focus solely on political identities – like partisan and ideological identity – but gain support among particular social groups(Mason 2016; Giles and Hertz 1994; Layman 1997). Moreover, during election campaigns like a presidential campaign, the candidates play the most crucial role in raising saliency of not only policy issues but more and more these social identities. Looking at the process of how social identities shape candidate preference based on the different identities the candidates activate is essential for three reasons. First, it helps us understand how important the processes of social sorting and mega-partisan identities are when it comes to candidate preference in a high-stake election. Second, we see which social identities (if any) are the most important for shaping candidate preference. Third, it indicates how well candidates' strategies in raising the salience of identities work and if they work expectedly. Moreover, particular interest lies in how social concepts that are not genuinely political influence political attitudes and, eventually political behavior. The next chapter will focus on the role of the saliency of social identities in U.S. Presidential Elections to develop an individual-level model to explain candidate preference in 2020 based on ethnic, national, and religious identity.

3.2. The Saliency of Social Identities in U.S. Presidential Elections

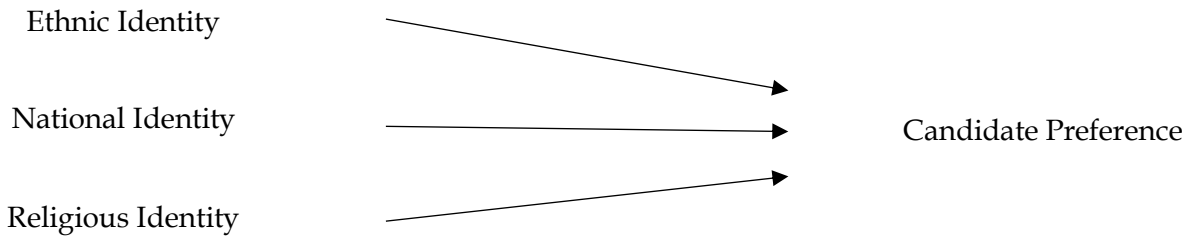
Saliency, or salience, is a concept with multiple differing definitions across scientific fields and even the same discipline (Boswijk and Coler 2020: 713). In the field of social psychology, Higgins (1996) defines a salient object as "information which is most easily brought to mind" (Higgins 1996: 133). Information can become salient because of certain stimuli or situations. However, salience differs from other related concepts such as accessibility, priming, or applicability. Salience can be conceptualized as not all features of a stimulus receiving equal attention at a single point in time (Higgins 1996: 135), while the other concepts focus on how quickly information is available, how it gets activated, and how well it fits a particular

situation. Higgins looks at salience as a characteristic of only one stimulus without the environment of potential other stimuli fighting for salience. However, this understanding does not seem very likely in a social situation. Individuals act and believe in social environments where they must decide which information is most relevant. Limited attention spans are central to human behavior in all parts of life (Falkinger 2008: 1596-97). Therefore, the salience of a concept like social identities is best understood in terms of comparative distinctiveness – an idea Higgins avoids in his work due to focusing only on a single stimulus (Higgins 1996: 136). Special attention is drawn to the differences between social groups, raising the salience of this particular group identity.

As previously mentioned, salience is an important concept regarding social identities. People only start acting on their group identity in situations that implement a clear cognitive "us" and "them" structure (Tajfel 1981: 287), and determinants such as similarity, equivalence, and interchangeability increase the salience and strength of a social identity even further (Turner 1987: 50). Therefore, if people find themselves in a situation that nourishes a cognitive "us" and "them" structure, highlights the similarities within groups, and distinguishes from other out-groups, people can be assumed to act and think primarily based on their group identity – the social identity becomes salient.

One of these special situations is a U.S Presidential Election. It is arguably the most important but certainly the most salient election in American politics. And since the contest is mainly between the two candidates of the two major parties, they are responsible for the topics becoming salient during the campaign. Candidates raising the salience of policy issues during an election campaign to increase their popularity is already well-studied (see Berliant and Konishi 2005; Abbe et al. 2003; Kiousis, Strömback, and McDevitt 2015). Since candidates can raise issue salience, it is only reasonable to assume that candidates have the power to raise social identity salience as well. If candidates successfully establish a cognitive "us" and "them" structure and thus activate and even strengthen specific identities, people could become more likely to act on their group identity. Therefore, people should prefer the candidate who activated and emphasized their identity. This results in a theoretical model in which ethnic, national, and religious identities affect candidate preference. Figure 1 is a graphical depiction of that model.

Figure 1: Directed Acyclic Graph for the Effects of Social Identities on Candidate Preference



Based on this model, one can now look at the situation of the U.S. Presidential Election 2020 to derive hypotheses for candidate preference. I will look at the two candidates – Joe Biden for the Democrats and Donald Trump for the Republicans – and how they raised the salience of certain ethnic, national, and religious identities. Starting with ethnic identity, Donald Trump was the candidate who strongly emphasized the white identity during the 2020 campaign. Kelly (2020) describes Trump's rhetoric as white ambivalence. The Republican candidate tried to speak to the white identity of people while not coming across as openly racist (Kelly 2020: 198). To achieve this, the incumbent president impoverished other ethnic identities like the black identity to frame the white identity as superior and capable of helping other ethnic groups to get out of their situation (Kelly 2020: 198). Additionally, he characterized other groups as lawless or uncivilized to highlight threats to the white identity (Kelly 2020: 203-04). In contrast, his challenger was appealing to ethnic identities other than the white identity. Biden was not creating a cognitive "us" and "them" structure. Instead, he was going for a more inclusive approach. This is seen in his focus on topics of racial justice and systemic racism during the 2020 campaign (Cineas 2020) as well as his focus on black voters and the legacy of the Civil Rights Movement (Hook and Mason 2020). Moreover, he played the counterpart to Trump on issues concerning immigration with proposals of raising annual ceilings for refugees and establishing a path to citizenship for undocumented immigrants already in the U.S. (Kaplan and Glueck 2019; Nagle 2019). Thus, Biden's focus on other ethnic groups and Trump's efforts to emphasize a threat to white identity should increase the salience of white identity so that people with a stronger white identity are more likely to prefer Donald Trump.

The picture of national identity looks very similar. In both elections, Donald Trump ran on the issue that other countries threaten the American people. He made other countries responsible for "bad" trade deals or income losses and defined them as a threat to American national security (Löfflmann 2022: 550-51). His appeal to the American national identity was based on "nostalgic reimagination" and "past national greatness" (Löfflmann 2022: 551), which materialized in his famous campaign slogan "Make America Great Again". Meanwhile, Joe Biden was focused on the possible benefits of international cooperation instead of the threats

other countries pose. Multilateralism and America's standing in the world were more important to the Democratic challenger (Lee and Weissert 2020). Donald Trump was the candidate primarily appealing to the national identity. A mixture of depicting other countries as threats and nostalgic reimagination created the environment necessary for national identity to become salient. Therefore, a stronger preference for Donald Trump can be expected among those with a stronger national identity.

Lastly, both candidates put different emphases on religious identity. Trump has a long history of speaking to primarily Christian people. He does it in two different ways. First, through the strong focus on Islamic terrorism in his speeches and framing Muslims as cultural enemies to the Christian Americans (Whitehead, Perry, and Baker 2018: 150). Second, Trump already used his 2016 campaign to stress that other religions around the world are attacking the American Christian heritage (Whitehead, Perry, and Baker 2018: 151). He again uses the threat to identity to increase the salience of Christian religious identity. On the other side, Biden was more restrained in mentioning religious groups and even his own religion. However, he used religious images such as hope, light, and faith in his major speeches (O'Loughlin 2020). Still, Trump tried to frame Biden as "against God" (Subramanian 2020). The cognitive "us" and "them" structure Trump creates in the context of religious identity compared to Biden's more subtle approach should lead to a stronger appeal of Donald Trump to people holding a religious Christian identity.

Three hypotheses derive from the salience structure created by the candidates in the U.S. Presidential Election 2020. Based on the theoretical model in Figure 1 built upon the salience of social identities, the hypotheses formulate relationships between the three identities and candidate preference in 2020.

> H_1: *The stronger the white identity, the higher the likelihood of preferring Donald Trump over Joe Biden.*
>
> H_2: *The stronger the national identity, the higher the likelihood of preferring Donald Trump over Joe Biden.*
>
> H_3: *People identifying as Evangelical or other Christians are more likely to prefer Donald Trump over Joe Biden than people identifying with another religion or none.*

Nonetheless, social identities are not the only predictors of candidate preference. Other theoretical approaches, such as the Economic Voting Theory have proven their explanatory power. The next chapter introduces sociotropic economic considerations as an alternative

explanation for candidate preference. This increases the strength of the argument of the importance of ethnic, national, and religious identity if the effects still hold when one of the most important predictors of political attitudes and behavior is included in the analysis.

3.3. Economic Considerations in Candidate Preference

Still, one of the most important explanatory approaches for political attitudes and behavior is the Economic Voting Theory (Lewis-Beck and Paldam 2000: 113-14). Especially in American electoral politics, economic considerations play a crucial role in decision-making (Gomez and Wilson 2001: 901). However, the literature is undecided on whether people act more in economic self-interest or public interest. The question of whether voters make decisions based on their pocketbook or sociotropic considerations has motivated scholars since the 1970s. This chapter lays down the main arguments and theoretical mechanisms of both approaches before eventually incorporating economic considerations into the existing individual-level model explaining candidate preference.

Before talking about the pocketbook-voting and sociotropic-voting hypotheses, the two main theoretical mechanisms of economic voting theory must be revisited. According to the economic voting approach, voters hold the government accountable for the economic situation – responsibility hypothesis (Lewis-Beck and Stegmaier 2019: 249). Only if voters connect the government to either their own economic situation or the country's situation do mechanisms of economic voting become important. They work in a way that governments get rewarded by voters for good economic situations while they get punished for bad economic situations – reward-punishment hypothesis (Lewis-Beck and Paldam 2000: 183). It raises the question, based on which considerations are these evaluations made?

The pocketbook-voting hypothesis states that voters assess candidates based on their own economic situation (Gomez and Wilson 2001: 899). The central assumption of pocketbook voting is that pocketbook voters are influenced mainly by the circumstances in their private lives. Since voters experience difficulties during periods of economic downturn, they turn to the polls to change the situation. If their personal economic situation is already very low, the perception emerges that any change, like a new President, will improve their situation (Kinder and Kiewiet 1981: 130-31). Additionally, all these evaluations come with relatively small information costs. Everybody should be aware of their own economic situation and can make a political decision based on this information. Using the own pocketbook as the main frame of reference for political attitudes and behavior seems to be something everybody can do.

In contrast, the sociotropic-voting hypothesis states that voters assess candidates based on the perceived health of the national economy (Gomez and Wilson 2001: 899). The sociotropic voter is influenced by the national economic condition or their perception of the national economy. The party in power suffers if the national economy is perceived as worse because voters blame the incumbents for the worse situation. And again, it does not take much information to participate in sociotropic voting. A rough evaluation of the national economy is sufficient to credit or blame an incumbent (Kinder and Kiewiet 1981: 132). However, if these evaluations are made based on altruism or self-interest cannot be determined (Kinder and Kiewiet 1981: 132).

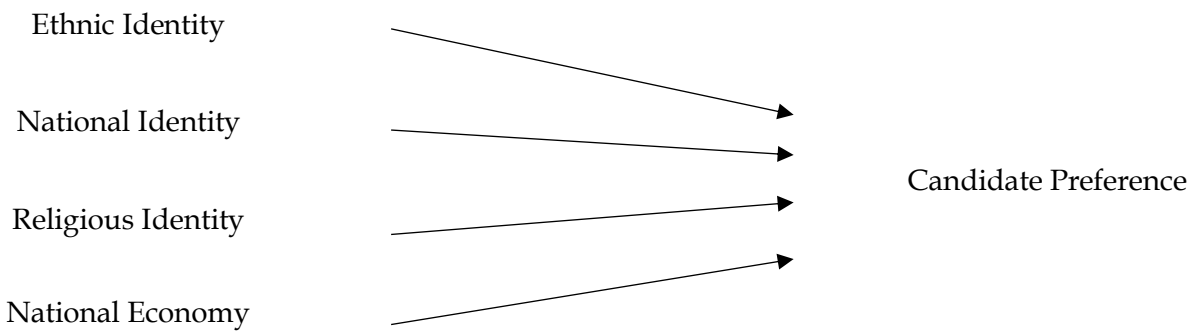
While studies find similar importance for the pocketbook-voting hypothesis and sociotropic-voting hypothesis (Healy, Persson, and Snowberg 2017: 783), most scholars give sociotropic considerations the edge in explaining political attitudes and behavior (Lewin 1991; Kinder and Kiewiet 1981; Lewis-Beck and Stegmaier 2019). In his attempt to determine whether voters act based on self-interest or public interest, Lewin finds no evidence for the pocketbook-voting hypothesis; instead, voters act based on what is best for the country (Lewin 1991: 45). And Lewis-Beck and Stegmaier provide evidence for an increased likelihood of favoring the incumbent when the perception of the national economy has changed from worse to better (Lewis-Beck and Stegmaier 2019: 252). Kinder and Kiewiet provide two explanations for the importance of sociotropic considerations in politics. On the one hand, they argue that the personal economic situation is more understood as a result of the immediate environment and personal failure. In contrast, only the national economic situation is projected onto the president and congress (Kinder and Kiewiet 1981: 157). On the other hand, they emphasize the different characters of information. Therefore, events of daily life are typically not politicized, while information on the national economy is genuinely politicized by sources of information like mass media (Kinder and Kiewiet 1981: 158).

Although economic considerations have a long tradition in the study of political attitudes and behavior, there are still limitations and pitfalls regarding the effects. First, Healy and colleagues criticize the temporal sequence. They argue that it is also possible for voters to first decide on their candidate or party preference and perceive the national economic situation in line with their preference (Healy, Persson, and Snowberg 2017: 774). Second, the perception of the national economy might be prone to partisan biases (Zaller 1992: 236-37). Voters might come to different evaluations of the national economy depending on which party is in office. Third, political sophistication might lead to heterogeneity in the effects of economic considerations. But while pocketbook-voting is a factor mostly among politically sophisticated

people, sociotropic considerations seem to be important for people of all sophistication levels (Gomez and Wilson 2001: 911).

Still, sociotropic considerations have shown to be an essential factor in political attitudes and behavior and should therefore be considered when looking at candidate preference in the U.S. Presidential Election 2020. Figure 2 shows the incorporation of the perception of the national economy into the individual-level model presented earlier. As a competing explanation for candidate preference to the social identity approach laid out earlier, sociotropic considerations allow seeing if the effects of social identities still hold if one of the strongest and most important predictors of political attitudes and behavior is considered.

Figure 2: Directed Acyclic Graph for the Effects of Social Identities and Perception of National Economy on Candidate Preference



The extension of the model leads to an additional hypothesis based on the Economic Voting Theory. In the U.S. Presidential Election 2020, Donald Trump was the Republican incumbent while Joe Biden was the Democratic challenger. In line with the reward-punishment hypothesis of Economic Voting Theory voters should decide on their candidate preference based on how they perceive the national economy's health. People perceiving the national economy as worsened should hold the incumbent responsible for the situation and therefore turn to the challenger. In the 2020 election, this is equal to people perceiving the national economy as worsened preferring Joe Biden over Donald Trump and vice versa.

> H_4: *People evaluating the national economy as worse have a higher likelihood of preferring Joe Biden over Donald Trump than people evaluating the national economy as better.*

Figure 2 now describes the whole theoretical model at the individual level. Sociotropic considerations join white identity, national identity, and religious identity in explaining candidate preference in the U.S. Presidential Election 2020. And H1 through H4 formulate hypotheses for the relationships based on the theoretical background. However, while various scholars have attempted to explain candidate preference with individual-level determinants,

this work will further extend the individual-level model with contextual-level factors. This will allow for a multilevel analysis that is still missing in the context of candidate preference.

3.4. Population Structure, Economy, and Political Context in the U.S. States

The theoretical model so far covers the individual-level determinants explaining candidate preference from a social identity perspective and with sociotropic considerations. Previous research has shown that the context in which political attitudes and behavior exist must be considered to get the full picture (André and Depauw 2017). Especially social networks on the county- and neighborhood-level matter for political preferences (MacKuen and Brown 1987). The homogeneity and density of social networks seem to determine social and political attitudes (Bienenstock, Bonacich, and Oliver 1990: 153). However, the state-level holds a special position in the election of the U.S. President. Due to the election process with the Electoral College, which electors from all 50 states constitute, and the winner-take-all principle in most states, the election has the character of multiple small elections at the state level.

Looking at state-level determinants is important for two reasons. First, given the winner-take-all systems, the effects of state-level determinants do not have to be very strong to affect the outcome of a whole election. Second, candidates and parties still think in the unit of states when they plan their campaigns and where they will spend the most money (Powell 2004: 121). For these two reasons, the large regional differences that come with different states in the U.S. will extend the theoretical model to a multilevel model of candidate preference in the U.S. Presidential Election 2020.

The analysis will focus on three factors at the contextual level: the population structure, the economic context, and the political context at the state level. Previous research has shown that these three contextual factors are important for the development of attitudes at the individual level. Oliver and Wong (2003: 580) argue that the heterogeneity of a population influences an individual's levels of prejudice and hostility toward other ethnic groups. And especially in an election featuring two candidates with very different positions on ethnic diversity as in the U.S. Presidential Election 2020 the ethnic diversity of an individual's environment might become important for candidate preference. The economic context is important in the case of the Presidential Election 2020 because of the contest between the incumbent and a challenger. The responsibility hypothesis of Economic Voting Theory suggests different candidate evaluations based on the economic context individuals live in because the incumbent is held accountable for the economic situation (Lewis-Beck and Stegmaier 2019: 249). Moreover, the

local political environment has a long tradition of influencing individual political attitudes (Putnam 1966: 640). However, the theoretical model will look at the state-level political context because of the above-mentioned special role of states in U.S. Presidential Elections.

The population structure is shaped by each state's composition of ethnic groups. This becomes important in the light of Allport's contact hypothesis formulated in 1954 because social identities lead to prejudice toward out-groups, which might shape political attitudes and behavior like candidate preference. The central premise of the contact hypothesis is that contact with out-group members exposes people to behavior that differs from the stereotype and therefore changes the perception of a group (Hamilton and Crump 2004: 483). However, this premise does not always work. Sometimes people do not change their perception of a group but instead create their own subgroups for people differing from the original stereotype (Hamilton and Crump 2004: 484). To have an effect, contact between different social groups must occur under certain conditions. Only if equal group status, common goals, intergroup cooperation, and support from authorities are given in inter-group contact, can it reduce prejudice (Allport 1954: 281). However, scholars also find that inter-group contact reduces prejudice if not all of the conditions are met (Mutz 2002: 113). There are four psychological mechanisms leading to inter-group contact reducing prejudice. Liebkind finds that continuous inter-group contact promotes learning about the out-group, changing one's own behavior, generating ties with members of other groups, and learning more about the in-group (Liebkind 2004: 345). All these effects reduce prejudice toward a specific out-group or out-groups in general. If you take regional entities like states with different levels of ethnic diversity in their population composition, the contact hypothesis's effect becomes apparent. States with higher ethnic diversity allow for more inter-group contact, reducing stereotypes and prejudice on average. This should shape the candidate preference in the U.S Presidential Election 2020 since Joe Biden and Donald Trump put different emphasis on ethnic groups. Therefore, ethnic diversity will be included as an indicator of population structure in the theoretical model to consider different compositions of the population.

Probably the most studied relationship in electoral research is that of economy and politics. Even Campbell and colleagues question if the economic situation might favor the incumbent (Campbell et al. 1960: 400). The claims made here are again based on the Economic Voting Theory introduced in the previous chapter. Due to the strong connection between candidate preference and actual vote choice in a highly personalized election, the effects Economic Voting Theory predicts can be transferred to candidate preference. As mentioned before, the responsibility hypothesis is the first key to economic voting theory. It states that voters hold the government accountable for the economic situation. The second key to economic voting

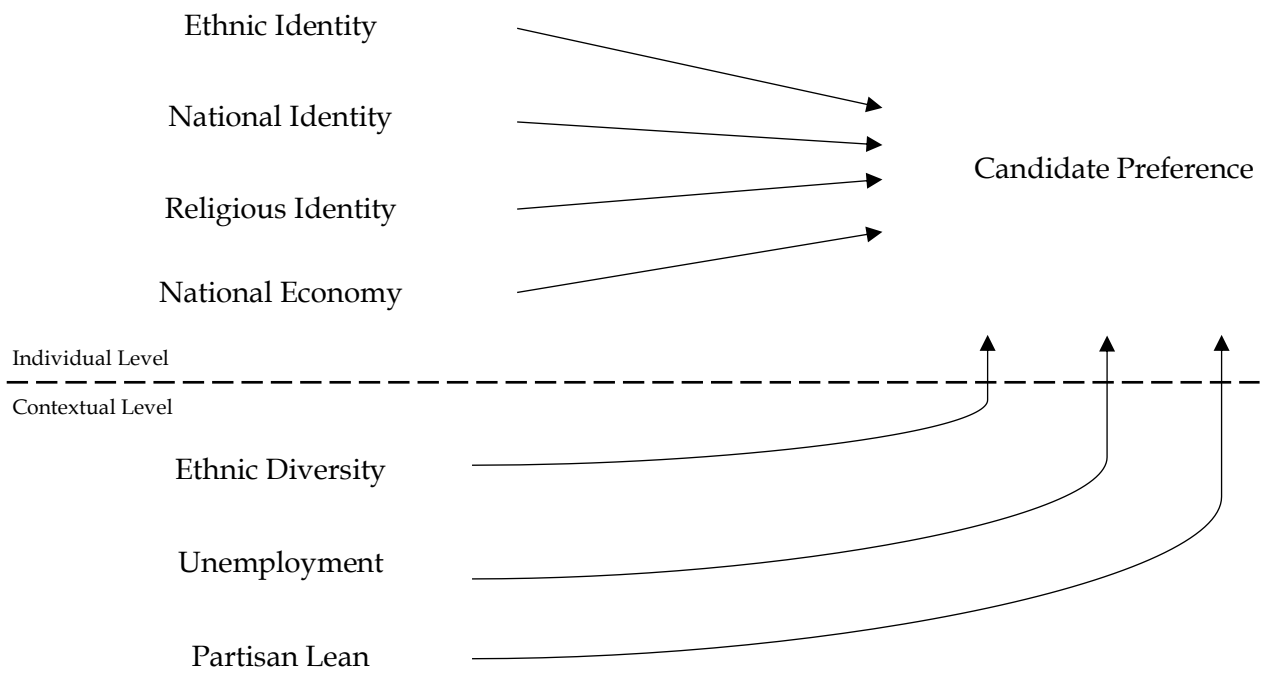
theory is the reward-punishment hypothesis. Voters reward governments for good economic conditions while punishing them for bad economic situations (Lewis-Beck and Paldam 2000: 183). The Economic Voting Theory is important in the context of the U.S. Presidential Election 2020 because of the contest between an incumbent, who is perceived to be responsible for the economic situation, and a challenger who represents a change from the current situation. Lewis-Beck and Stegmaier identify four generalizations that result from the Economic Voting Theory (Lewis-Beck and Stegmaier 2019: 251-52). First, sociotropic economic voting is more important than pocketbook voting. Second, the responsibility for the economic situation must be clear to the voter. Third, aggregate- and individual-level analyses of the economic situations came to the same conclusions. Fourth, unemployment and growth are the best indicators of the economic situation. The unemployment rate will be included in the theoretical model as an indicator of the economic situation in a state.

To say that the political context matters for political attitudes and behavior is almost self-evident, the same goes for candidate preference in U.S. Presidential Elections. The most discussed parts of political contexts in terms of state-level differences are the party affiliation of the governor and the partisan lean in a state. In 2020, 23 states were governed by Democratic governors while 27 were governed by a Republican. However, the governor's party is not necessarily the dominant party in a state. Looking at the influence a governor might have on candidate preference, most political commentators assumed an advantage for the presidential candidate of the same party as the governor (Erikson, Folke, and Snyder Jr. 2015: 491). Instead, Erikson and colleagues find the exact opposite to be true. Having a governor of the same party hurts the presidential candidate. The authors attribute this result to the ideological balancing effect: "Centrist voters seek a balance between the ideological excesses of the left and right" (Erikson, Folke, and Snyder Jr. 2015: 492). The main idea is that voters perceive the ultimate policy outcome as a mixture of national and subnational policies. Therefore, having a governor of the same party as the presidential candidate should hurt the presidential candidate in this state. However, evidence suggests that solely looking at governors does not capture the whole complexity of a political context in a state. Jacobson (2006) argues that the standing of governors in a state is based on cross-party approval, which a candidate needs to get elected governor in the first place (Jacobson 2006: 749). Presidential candidates do not necessarily need this cross-party approval in a state since they can rely on other states. This results in presidential candidate preference being more prone to actual partisan lean in a state than governors. Thus, the partisan lean is a more accurate indicator of the political environment come a Presidential Election. The partisan lean in a state should affect the preference for a candidate in terms of favorability of the candidate from the party the whole state is leaning

toward. Partisan lean will be included as indicator for political context in the theoretical model since it is the most accurate.

A look at the population structure, economy, and political context has shown that contextual-level characteristics must be considered to get the complete picture of candidate preference in the 2020 U.S. Presidential Election. Figure 3 enhances the theoretical for the individual-level features and includes the population structure, unemployment rate, and partisan lean as contextual-level determinants of candidate preference.

Figure 3: Directed Acyclic Graph for the Effects of Individual-Level and Contextual-Level Predictors on Candidate Preference



The enhancement of the individual level to a multilevel theoretical model comes with three additional hypotheses formulating expectations about the effects of the contextual-level determinants on candidate preference. First, if you take regional entities like states with different levels of ethnic diversity in their population structure, the contact hypothesis' effect becomes apparent. States with higher ethnic diversity allow for more inter-group contact, reducing stereotypes and prejudice on average. This gives an advantage to the more liberal candidate in this state regarding candidate preference. H_5 hypothesizes this relationship for the contest between Joe Biden and Donald Trump in 2020. Second, the Economic Voting Theory becomes important because of the president's responsibility for the economic situation perceived by the American people. Since Donald Trump ran for President as the incumbent in 2020, the economic situation should shape the people's perception of him, according to the theory. In good economic situations, he should be rewarded with more preference; in bad

economic situations, he should be punished. Since the unemployment rate is used as an indicator of the economic situation in a state in 2020, H_6 expects a lower preference for Donald Trump and a higher preference for Joe Biden in states with higher unemployment rates. Third, the effect of partisan lean in a state is almost self-explaining. Since Joe Biden was the candidate for the Democratic Party, he should perform better in states leaning more toward the Democrats while Donald Trump should perform better in states leaning more toward the Republicans. Therefore, people living in a state leaning toward the party of the candidate should prefer the party's candidate over the other candidate. H_7 describes this relationship in the context of the U.S. Presidential Election 2020.

> H_5: *The higher the ethnic diversity in a state, the higher the likelihood of preferring Joe Biden over Donald Trump.*
>
> H_6: *The higher the unemployment rate in a state, the higher the likelihood of preferring Joe Biden over Donald Trump.*
>
> H_7: *People living in a Democratic-leaning state have a higher likelihood of preferring Joe Biden over Donald Trump than people living in a Republican-leaning state.*

The next chapter will introduce the data and method used to test the theoretical model and its hypotheses. Data was gathered from various sources to allow for a multilevel analysis of candidate preference in the U.S. Presidential Election 2020 including all factors of the individual and contextual level of the theoretical model.

4. Data and Method

4.1. Sources of Individual- and Contextual-Level Data

Four data sources are used for the analysis. The American National Election Study 2020 (American National Election Studies 2021) is used for the individual-level data. The contextual-level data consist of state-aggregated U.S. Census 2020 data, unemployment data from the U.S. Bureau of Labor Statistics, and the state-level partisan lean from FiveThirtyEight. This section describes every data source and the variables used for the analysis.

All individual-level data is used from the American National Election Study survey. The ANES 2020 Time Series Study Full Release results from a random sample drawn from the United States Postal Service Computerized Delivery Sequence File, with every address having the same probability of being selected. The target population is 231 million U.S. citizens aged 18 years and older living in the 50 states or the District of Columbia. However, only one respondent from every household is interviewed, leading to an unequal selection probability for every individual. Post-stratified survey weights account for this unequal probability of selection in the dataset. The dataset consists of a pre- and post-election sample. The pre-election survey was conducted with 8.280 participants from August 18, 2020, to November 3, 2020. Of these 8.280 individuals, 7.449 agreed to a second post-election survey from November 8, 2020, to January 4, 2021. The survey modes were a web-only survey, a web survey with additional phone contact for non-responding cases, and a video interview with additional web survey or phone contact for non-respondents. Since variables from the post-election survey are used, the analysis will work with the post-election sample of 7.449 respondents. The provided survey weights for the post-election sample will be used to weigh the data.

Contextual-level data for the U.S. states in 2020 are gathered from three sources. First is the 2020 United States Census. It was the decennial count of the U.S. population in the year 2020. The census covers topics from demographic characteristics over health and education to the economic situation. For the analysis, state-level aggregated data is used to include information on the population structure at the state level in the modeling process. The United States Census Bureau (United States Census Bureau 2022) has released the aggregation of original census data. Using the census data allows the inclusion of accurate population data from the same year as the U.S. Presidential Election 2020. Second is the United States Bureau of Labor Statistics. This is the only official source for all states to get accurate unemployment data. To include unemployment data at the state level, the annual average unemployment rates by states for the year 2020 are used (U.S. Bureau of Labor Statistics 2022). Third is the opinion poll

website FiveThirtyEight. Although partisan lean at the state level is a measure calculated by multiple institutions like the Cook Partisan Voting Index, FiveThirtyEight is the only source providing historical data on the year 2020 while allowing for a complete understanding of how the index was composed. FiveThirtyEight's partisan lean measure is used to include data on the political context (FiveThirtyEight 2020).

Theoretical and practical reasons determine the eventual selection of cases. At the individual level, a subsample of people identifying as white in the ANES 2020 dataset is used due to data constraints. The number of observations of all other ethnic groups in the dataset is too small to include in the analysis. Therefore, all results are regarding people identifying as white. Moreover, all cases with missing values on either individual-level predictor variables are excluded listwise, and cases without information on which state the respondent is living in, resulting in a final sample size of 4.474 from the original 7.449. At the contextual level, all 50 U.S. states are part of the analysis.[1]

4.2. Dependent, Independent, and Control Variables

The dependent variable is the candidate preference between Donald Trump and Joe Biden in the U.S. Presidential Election. In the ANES 2020 survey, respondents were asked to rate both candidates on a scale from 0 to 100 degrees, like a thermometer, with 0 being very negative and 100 being very positive. The difference in candidate ratings is the size of the gap between those ratings for Joe Biden and Donald Trump. Respondents with a gap of fewer than 15 points are usually considered swing voters (Skelley 2019), and respondents with a gap of 15 points or higher are considered to prefer one candidate over the other. Favoring Joe Biden is coded as 1, and favoring Donald Trump is coded as 0. For example, if a respondent evaluates Joe Biden with 75 degrees and Donald Trump with 50 degrees, the difference would be 25 points. Therefore, this case would be coded as preferring Joe Biden (1). Respondents with missing values on at least one of the thermometer scales and respondents with no clear preference (evaluation within 15 points) are excluded from the analysis.

The independent variables consist of three categories of variables. First, white identity, national identity, and religious identity at the individual level from the ANES 2020 data are included. Second, the perceived health of the national economy from the ANES 2020 data is

[1] A table of all states included in the analyses as second-level clusters and the number of individual observations in each state can be found in Appendix 1.

retained as sociotropic consideration[2]. Third, ethnic diversity, unemployment rate, and partisan lean at the contextual level of states are included. All missing data was excluded listwise.

White identity and national identity were raised by the question: "How important is being white (or American) to your identity?". The scale ranges from 1 "Extremely important" to 5 "Not at all important". For a more intuitive interpretation of the results, the variables were reverse coded to lower values corresponding to lower importance and higher values corresponding to higher importance. Religious identity is measured by the question of which religion respondents identify with, if any. The nine possible categories are condensed into four theoretically more relevant categories of "Evangelical Christian", "Other Christian", "Other Religion", and "Not Religious". Therefore, white identity and national identity will be used as continuous variables and religious identity will be used as a categorical variable.

The perception of the nation's economic health is measured by the question: "Now thinking about the economy in the country as a whole, would you say that over the past year the nation's economy has gotten better, stayed about the same, or gotten worse?". The sociotropic economic evaluation will be included in the model as a categorical variable varying between "Gotten Better", "Stayed about the same", "Gotten Worse".

Ethnic diversity, unemployment rate, and partisan lean are the variables at the state level included in the analysis. Ethnic Diversity is calculated from shares of racial and ethnic groups in every state based on 2020 U.S. Census data. The ethnic diversity index ranges from 0 to 1. It represents the "probability that two people chosen at random will be from different racial and ethnic groups" (United States Census Bureau 2021), with 0 meaning no ethnic diversity and 1 meaning that every individual is of a different racial and ethnic group[3]. The original ethnic diversity index has been transformed to a range from 0 to 100 for better interpretation. The unemployment rate will be included for every state as the annual average of unemployment rates in 2020. The U.S. Bureau of Labor Statistics officially calculates the unemployment rate as the unemployed share of the labor force population, which is the sum of employed and unemployed people (U.S. Bureau of Labor Statistics 2022). Partisan lean by states in 2020 is calculated by FiveThirtyEight as "the average margin difference between how a state or district votes and how the country votes overall" (Rakich 2021). The partisan lean score of a state is based on results from previous elections. In the case of partisan lean in 2020, the U.S

[2] A table with full question wording and answering scales for the original variables from which the dependent variable has been created and all independent variables in the survey can be found in Appendix 2.
[3] The formula to calculate the ethnic diversity index can be found in Appendix 3.

Presidential Election 2016 is weighted at 50%, the U.S. Presidential Election 2012 is weighted at 25%, and the last four elections to the state legislature are weighted at 25% (FiveThirtyEight 2020). This results in a continuous measure of partisan lean, indicating how many percentage points a state leans differently from the nation. For a more intuitive interpretation of the results, the partisan lean measure was coded as a categorical variable only indicating if a state leans more toward the Democratic Party, the Republican Party, or neither.

In addition to the individual-level and contextual-level predictors, control variables are included in the model. Control variables in a regression framework aim to control for possible confounding effects which distort the relationship between independent and dependent variables (Spector and Brannick 2011: 288). In contrast to experimental studies, observational studies are prone to omitted variable bias (Wilms et al. 2021: 1), which means that one or more independent variables that affect the independent variables of interest and the dependent variable are not considered. To address this problem, it is common practice to include control variables that might be important for the causal relationships in the model on a theoretical basis. While some authors argue that control variables are often not used correctly and must not be used only because a majority of researchers use the technique of controlling (Spector and Brannick 2011: 288), their importance in observational studies remains uncontested if used correctly.

The independent variables of interest in the analysis are the white, national, and religious identities. To obtain unbiased estimates for these three variables, the effects on candidate preference are controlled for education and gender. Despite other demographic variables like age being often used in the research of political attitudes, the existing literature does not give a theoretical explanation for its importance in the relationship between social identities and candidate preference.

First, the model estimates will be controlled for educational level. Recent research has shown the importance of education when it comes to political preferences and social identities. All sophistication levels – if college degree or not – get overall information on the economy (Enns and Kellstedt 2008: 450) or major political developments (Zingher and Flynn 2019: 132). Therefore, they might differ in their preferences in a politically highly salient situation such as the presidential election. Moreover, as Zingher and Flynn point out, less sophisticated individuals also play a crucial role in social sorting (Zingher and Flynn 2019: 144). Since social sorting happens along identities such as ethnic, national, or religious identity (Mason 2015: 131), an influence of the educational level on the development of social identities can be assumed. Hence, one should control for education when looking at the relationship between

social identities and candidate preference. Second, the model controls for possible gender differences. Norrander and Wilcox (2008) find differences between men and women in their issue-orientation and ideological and religious identity. While men and women in America have moved to the conservative side of the self-placement scale, a growing number of women identify as more liberal. Especially the well-educated, single, and non-evangelical woman is set to be more liberal (Norrander and Wilcox 2008: 521). Additionally, women are more likely to base their ideological identification on gender-specific issues (Norrander and Wilcox 2008: 521). Therefore, gender should be included as a control variable since it has important implications on how people build their attitudes and identities.

Nonetheless, the estimates of education and gender cannot be interpreted in the same way as the estimates of the identity variables. Westreich and Greenland argue that the estimates of the control variables and the independent variables of original interest represent different causal effects (Westreich and Greenland 2012: 293). Since the effects of the identity variables are controlled for education and gender, they can be interpreted as the total effect of identity on candidate preference conditional on the control variables. The total effect is defined as "net of all associations of a variable through all causal pathways to the outcome" (Westreich and Greenland 2012: 292). However, the effects of the control variables can only be interpreted as the direct effects, which are not mediated through the control variables' effects on the identity variables (Westreich and Greenland 2012: 293). In other words, while one tries to obtain an unbiased estimate for the identity variables and interprets the estimates as such, this is not the case for the control variables. To avoid this "Table 2 Fallacy" – as Westreich and Greenland call it – the control variables will be included in the model, but the effect estimates will not be displayed in the results table, as recommended by the authors (Westreich and Greenland 2012: 296).

4.3. The Random-Intercept Model

Analyzing observational data like the American National Election Study 2020 allows for using some form of regression framework to test hypotheses on relationships between the independent and dependent variables. Depending on whether the outcome variable is continuous or categorical simple ordinary least squares or logit regressions are typically used for analyses. However, a series of assumptions must be met for these models to produce robust results. One is the assumption of independently distributed error terms for individual observations. This means there must not be a relationship between individuals – or serial correlation. While serial correlation does not automatically produce biased coefficients, it leads to wrong estimations of standard errors and, therefore, wrong results in terms of statistical

significance (Finch, Bolin, and Kelley 2019: 29; Hox 2010: 5). The null hypothesis gets rejected more often than expected although it is true in the population. Serial correlation plays an important role if data is spatially or temporally structured (Ziller 2018: 468). Spatially structured data exists if individual observations are clustered within a geographical region. For example, these data structures can be found in cross-country surveys with individuals clustered in different countries. In this case, the error terms of individual observations (people) are not independently distributed but correlate within clusters (countries). Temporally structured data is created with panel surveys asking the same respondents at different times. In this case, the error terms are correlated within one respondent. Serial correlation in hierarchical data makes it difficult to perform simple regression analyses. More sophisticated multilevel models account for serial correlation in hierarchical data by including variables for the second level (cluster-level).

The multilevel regression model is a family of modeling strategies that accounts for different levels within the data. The advantages and disadvantages of different multilevel models will be considered to select the suitable model for the analysis. When accounting for hierarchical data, one can choose between fixed-effects and random-effects models. Both control for correlated error terms between individual observations in different ways. While fixed-effects models include dummy variables for the group-level, random-effects models include additional error terms to allow for varying effects across groups (Ziller 2018: 468).

As mentioned before, the fixed-effects model includes a dummy variable for every group at the second level. Through this method, the fixed-effects model controls for all differences that might exist between the groups. Therefore, it comes with the advantage of controlling for all unobserved heterogeneity that might exist at the group level (Ziller 2018: 479). Moreover, the fixed-effects model is less complex than a random-effects model since it does not include additional error terms for every group. However, despite all the advantages of the fixed-effects model, it does not allow for having explanatory variables at the group level. Therefore, the fixed-effects model is not suited to examine the effects of theoretically relevant group-level variables on the individual-level outcome. If the research interest lies in the effects of group-level variables, one must turn to random-effects models.

The simplest form of random-effects models includes an additional error term for the group level, allowing the intercept to vary across different groups. That is, the model allows for different means of the outcome variable in every group if all independent variables are equal to 0 (Finch, Bolin, and Kelley 2019: 30). While the outcome must be on the lowest level, independent variables can be included from every level. In the example of cross-country

surveys, independent variables can be at the individual level (e.g., political interest) and the contextual level of countries (e.g., a country's GDP). The random-intercept model not only allows for group-specific intercepts and the inclusion of independent variables from different levels but also for interactions between level-1 and level-2 predictors. Cross-level interactions can be used to find group-level predictors that moderate the effects of individual-level predictors on the outcome variable. Random-effects models even allow for coefficients to vary across groups in a more sophisticated approach. This random-coefficient (or random-slope) model includes further error terms for every coefficient of interest. The idea behind random-coefficient models is that the effects of explanatory variables on the dependent variable differ across groups. However, random-effects models run into a variety of problems. The model can become very complex depending on the number of different levels and predictors included. Moreover, since an error term for every group is included instead of dummy variables, as in the fixed-effects model, the random-effects model does not account for all potential unobserved heterogeneity at the group level.

The research interest of this thesis lies in the effects of predictors at the individual level and the state level on candidate preference in the U.S. Presidential Election 2020. Therefore, the random-effects model is chosen for analysis because of the advantages these types of models provide. A fixed-effects model would not allow for the inclusion of predictors at the group level. The problem of potential unobserved heterogeneity remains as in every non-experimental study and will be considered when interpreting the results. Moreover, the analysis will only consist of a random-intercept model for two reasons. First, there are no theoretical expectations for the effects of the independent variables to vary across states. Second, including random slopes would increase the complexity of the model to the extent that is out of this work's scope.

Before applying a multilevel analysis in form of a random-intercept model, one must determine whether the data is hierarchical. There are two ways to decide whether a multilevel analysis is necessary or not. First, multilevel analysis is necessary if the sampling strategy indicates non-independent observations (Finch, Bolin, and Kelley 2019: 23). This is the case in most cross-country surveys that select random samples from every country and compile this into one extensive dataset. Observations from the same country are often not independent from one another. Second, multilevel analysis is necessary if there is a certain proportion of variation in the outcome at the group level (Finch, Bolin, and Kelley 2019: 28). The frequently used method to determine the percentage of variation at the group level is calculating the intra-class correlation coefficient (ICC). Although there is no universally agreed upon threshold, LeBreton and Senter already attribute a small to medium effect to the group level if

the ICC is at least 0.05 (LeBreton and Senter 2008: 838). Calculating the ICC for candidate preference between Joe Biden and Donald Trump in the U.S. Presidential Election 2020 with the state as a grouping factor results in an ICC of 0.089. Therefore, 8.9% of the variation in the dependent variable can be attributed to the state level, which allows for multilevel analysis.

4.4. Research Strategy

As described in the data section, the outcome variable of candidate preference between Joe Biden and Donald Trump in the U.S. Presidential Election 2020 is a binary outcome. For binary outcomes, the assumption of normality of the dependent variable is violated, which has to be accounted for (Hox 2010: 112). Multilevel Generalized Linear Models control for non-normally distributed outcomes and heteroscedastic errors by using a link function with its corresponding error distribution. Every Multilevel Generalized Linear Model has three distinct components: a link function, a specific error distribution, and a multilevel linear regression equation (Hox 2010: 121). The most frequently used link function for binary outcomes is the logit function. The logit link function creates a linear model for the logarithmic odds of the binary outcome variable, taking a value of 1 (Rabe-Hesketh and Skrondal 2012b: 503). The advantage of using a logit function is the simple linear regression framework on the right side of the equation. This allows for the inclusion of different types and combinations of variables as in linear regression (Rabe-Hesketh and Skrondal 2012b: 502). However, coefficients resulting from a Multilevel Generalized Linear Model using a logit function can only be interpreted regarding the direction of effects. For more substantial interpretations of effects, one must use the change in odds or predicted probabilities which can be obtained by transforming the given logit coefficients. The corresponding error distribution for dichotomous outcome variables in Multilevel Generalized Linear Models with a logit function is the Bernoulli distribution with a mean value of μ. Bernoulli is a particular case of binomial distributions with only one trial with an outcome of 0 or 1 instead of a proportion of outcomes in multiple trials (Hox 2010: 122).

A Multilevel Generalized Linear Model with a logit link function – or Multilevel Logistic Regression – is used to calculate a random intercept model for candidate preference in the U.S. Presidential Election 2020. The analysis is conducted with the statistical programming language R. Equation 1 states that the binary outcome of candidate preference for every individual (i) in a group (j) is used as a proportion (π_{ij}) in the Multilevel Logistic Regression assuming a Bernoulli distribution with mean μ. π_{ij} will be predicted via a logistic regression, as stated in Equation 2. The random intercept model is represented in parentheses. The equation consists of fixed and random parts. The average intercept (or grand mean) stays the

same for all groups and is, therefore, a fixed effect displayed as γ_{00}. This is the predicted value for our outcome if all independent variables are equal to 0. The random part of the model is the group-specific effect on the average intercept U_{0j}. U_{0j} is the deviation between the average intercept and the group-specific effects on this intercept. Since the random intercept model only allows the intercept to vary across groups, all other effects in Equation 2 are fixed effects. γ_{10} to γ_{70} are the coefficients of the individual level predictors, while γ_{01} to γ_{03} are the coefficients of the contextual level predictors in the model.

(Equation 1)

$$candidate\ preference_{ij} = \pi_{ij}\ ;\ \pi \sim Bernoulli(\mu)$$

(Equation 2)

$$\begin{aligned}\pi_{ij} = logistic\big(&\gamma_{00} + \gamma_{10} white\ identity_{ij} + \gamma_{20} national\ identity_{ij} + \gamma_{30} religious\ identity_{ij} \\ &+ \gamma_{40} national\ economy_{ij} + \gamma_{50} gender_{ij} + \gamma_{60} education_{ij} \\ &+ \gamma_{01} ethnic\ diversity_j + \gamma_{02} unemployment_j + \gamma_{03} partisan\ lean_j + U_{0j}\big)\end{aligned}$$

The Multilevel Logistic Regression is estimated by Maximum Likelihood Estimation. This estimation method is an iterative process to estimate population parameters. It calculates the model with the highest likelihood of obtaining the observed data (Finch, Bolin, and Kelley 2019: 36). The model estimation takes place via a comparison of observed data with predicted data and stops at the point when the difference between observed and predicted values do not get smaller – or in other words when the model converges (Hox 2010: 41). However, Maximum Likelihood Estimation does not account for the number of parameters being estimated when determining degrees of freedom which can lead to underestimation of the model's variance components. Restricted Maximum Likelihood controls for this problem by including the number of parameters estimated when determining degrees of freedom (Finch, Bolin, and Kelley 2019: 36). Still, when running a multilevel model with a large number of groups at the second level, the differences between Maximum Likelihood and Restricted Maximum Likelihood estimations become very small (Finch, Bolin, and Kelley 2019: 36-37).

All continuous explanatory variables are standardized for the multilevel analysis. Centering or standardizing independent variables are often used as transformation techniques to make 0 a meaningful value. This simplifies the interpretation of the multilevel logistic regression results since the intercept is the value of the predicted logit if all independent variables are equal to 0 (Finch, Bolin, and Kelley 2019: 34). While centering subtracts the mean value of a

variable from every observed value resulting in a mean of 0, standardizing subtracts the mean and divides by the standard deviation, which results in a mean of 0 and a standard deviation of 1. Standardizing explanatory variables also changes the units of interpretation. The coefficients are not interpreted as changes in the predicted logit when the independent variable increases by one unit of the original scale anymore. Instead, the new unit of interpretation is the standard deviation. The multilevel model's standardized coefficients of continuous variables must be interpreted as changes in the predicted logit if the independent variable is increased by one standard deviation.

The multilevel analysis is conducted in five steps. First, the null model is included to get baseline values of intra-class correlation and R^2. Second, all three identity variables will be included at the individual level. The identity model provides insight into the effects of white, national, and religious identity on candidate preference. Third, sociotropic economic considerations and control variables are added in a third model. Including the perceived health of the national economy will show if the effects of the identity model still hold if one of the most important predictors of political attitudes is considered. Fourth, the contextual-level determinants of ethnic diversity, unemployment, and partisan lean of all 50 U.S. states are included. The full model gives information on the effects of contextual-level determinants on candidate preference and possible changes in the effects of individual-level variables. In the fifth and last step, further analyses may be conducted on the effects that stand out the most regarding size and direction. The analysis will close with contributions to the field of research and limitations regarding data constraints, research design, and causality.

Like other regressions, multilevel regression comes with a variety of assumptions that must be true for the results to be statistically valid. Model diagnostics have been performed for the full multilevel model including individual-level and contextual-level determinants. According to Rabe-Hesketh and Skrondal (2012a: 160-61), normally distributed level-1 residuals and normally distributed random intercepts are assumed for multilevel analysis. Since the analysis does not include random slopes, no test for normally distributed random slopes was performed. The inspection of the distributions of the level-1 residuals and the random intercepts showed only slight deviations from a normal distribution.[4] Therefore, the analysis can be performed without any limitations due to the violation of assumptions. The following chapter presents the results of the analysis.

[4] Histograms containing the distribution of level-1 residuals and random intercepts can be found in Appendix 4.

5. Results

The candidate preference in the U.S. Presidential Election 2020 is analyzed from different perspectives to assess the hypotheses formulated in chapter 3 and answer the research questions. First is a univariate inspection of the dependent variable and independent variables. Second, bivariate relationships between each independent variable and the dependent variable of candidate preference provide further insights and allow for a first assessment of the hypotheses. Third, multilevel logistic regression is used to fully evaluate the hypotheses and look at effect sizes and directions.

5.1. Univariate and Bivariate Analyses

The univariate tables contain weighted data for the individual-level variables used from the ANES 2020 dataset. Since variables from both parts of the survey (pre-election and post-election sample) are included, the full sample post-election weight provided in the dataset was used to weight the data. Table 1 shows the summary statistics for the original variables for the evaluation of Joe Biden and Donald Trump on a continuous scale from 0 to 100. The categorical dependent variable for the analysis was created out of these two variables. Additionally, the weighted frequencies for the categorical outcome are included. Since the frequencies are weighted, the sum does not add up to the same number of observations included in the analysis (n = 4474).

Table 1: Weighted Summary Statistics and Frequencies of Candidate Preference

CONTINUOUS VARIABLES: ORIGINAL VARIABLES		**MEAN**	**MEDIAN**	**MIN.**	**MAX.**	**STD. DEVIATION**	**VARIANCE**	**N**
	Biden Thermometer	*42.13*	*40*	*0*	*100*	*35.02*	*1226.57*	*4474*
	Trump Thermometer	*48.00*	*60*	*0*	*100*	*42.35*	*1793.25*	*4474*

CATEGORICAL VARIABLE: CANDIDATE PREFERENCE		**FREQ.**
	Joe Biden	*1776*
	Donald Trump	*2024*

The summary statistics of the continuous variables show that Donald Trump was evaluated higher on average with a mean value of 48 compared to a mean evaluation of around 42 for Joe Biden. The same goes for the median values of 40 for Joe Biden and 60 for Donald Trump.

Moreover, the standard deviations indicate that the values for Joe Biden are more clustered around the mean than those for Donald Trump. This is in line with the frequencies of the categorical candidate preference variable. The original means indicate a better evaluation on average of Donald Trump. More respondents prefer Donald Trump over Joe Biden, with a difference of at least 15 points on the continuous scale.

Table 2 displays the weighted summary statistics for all continuous independent variables and frequencies for all categorical independent variables at the individual level. Included are the three identity variables (white identity, national identity, religious identity), the perception of the national economy, and the control variables (education, gender). The white identity's mean of 2.28 and the national identity's mean of 3.75 show higher importance of being American to the respondents' identities than being white. The standard deviations indicate similar variation around the mean for both variables. The frequencies for the religious identity variable illustrate that most respondents identify as Christian other than Evangelical and least identify with another religion. Identifying as Evangelical or as not religious are in between. Moreover, the majority perceived the national economy has worsened over the past year, while a comparable amount perceived it has gotten better or stayed about the same. The sample used in the analysis consists of 1949 women and 1851 men. And 2039 of the respondents earned a college degree while 1762 did not.

Table 2: Weighted Summary Statistics and Frequencies of Individual-Level Independent Variables

CONTINUOUS VARIABLES		MEAN	MEDIAN	MIN.	MAX.	STD. DEVIATION	VARIANCE	N
	White Identity	*2.28*	*2*	*1*	*5*	*1.27*	*1.61*	*4474*
	National Identity	*3.75*	*4*	*1*	*5*	*1.23*	*1.52*	*4474*

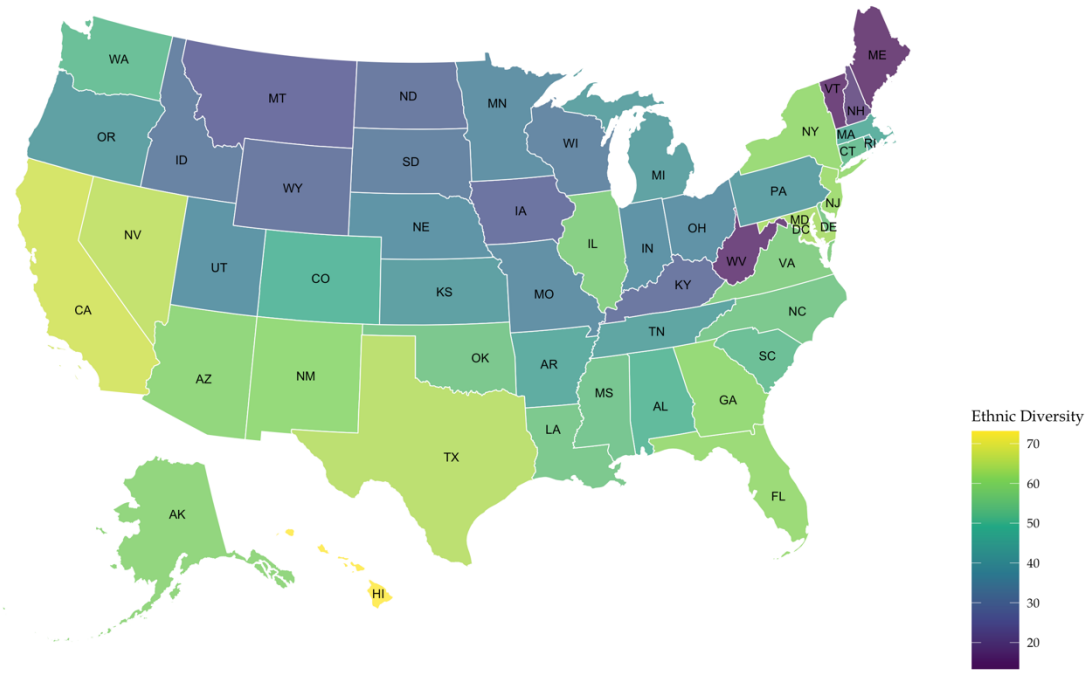
CATEGORICAL VARIABLES	RELIGIOUS IDENTITY	FREQ.	NATIONAL ECONOMY	FREQ.	GENDER	FREQ.	EDUCATION	FREQ.
	Evangelical Christians	*624*	Gotten Better	*834*	Female	*1949*	College Degree	*2039*
	Other Christians	*2155*	Gotten Worse	*2162*	Male	*1851*	No College Degree	*1762*
	Other Religion	*188*	Stayed about the same	*804*				
	Not Religious	*833*						

The contextual level variables are visualized in the Figures 4 to 6. The figures show maps of all 50 U.S. states colored by ethnic diversity, unemployment, and partisan lean of each state in

the year 2020. For convenience, the states of Alaska and Hawaii are depicted below the map of the other 48 states. All states are labeled with their two-letter abbreviations. Legends are displayed on the right side of the maps.

Figure 4 shows the Ethnic Diversity Index score for every state. The Ethnic Diversity Index ranges from 0 to 1, indicating how ethnically diverse a state's population is. A value of 0 corresponds to complete ethnic equality, while a value of 1 corresponds to complete ethnic diversity. Darker colors in Figure 4 represent low ethnic diversity, while brighter colors represent higher ethnic diversity. Ethnic diversity in the U.S. is lowest in the North-Eastern region, including states like Maine, New Hampshire, and Vermont. Moreover, the Midwestern states and states with low population sizes, like Wyoming and Montana, are more likely to have little ethnic diversity. In contrast, the East Coast, West Coast, and Southern states are the regions with the highest scores of ethnic diversity. The average value of ethnic diversity is 0.45, with Hawaii having the highest (0.73) and Maine having the lowest (0.13) scores.

Figure 4: Ethnic Diversity Index by State in 2020 (Source: Own Representation)



The unemployment rate is displayed in Figure 5. The unemployment rate by state is the share of unemployed people in the labor force. Like Figure 4, brighter colors indicate higher values, and darker colors indicate lower values. Nebraska was the state with the lowest unemployment rate in 2020 with 4.1%, while Nevada had the highest unemployment with

13.5%. The average U.S. state had an unemployment rate of 7.4% in 2020. However, Figure 5 shows that the state of Nevada was an outlier in 2020. This can be explained by the exceptionally high unemployment rates in tourism-related businesses, which suffered particularly during the beginning of the COVID-19 pandemic in early 2020 (Kelly and Morell 2020). Moreover, the states on the West Coast and in the Rust Belt region were among those with higher unemployment rates. In contrast, more rural states in the Midwest came with the lowest unemployment in 2020.

Lastly, Figure 6 shows the partisan lean of each state. Partisan lean originally means by how many percentage points a state leans more toward one party compared to the nation. Figure 6 displays the recoded categorical variable of partisan lean, showing whether the state leans toward the Democrats (blue) or the Republicans (red). Some Midwestern states join the East Coast and West Coast in leaning toward the Democratic Party, while most states (31) lean toward the Republicans. Especially the Southern and Midwestern States are dominated by the Republican Party. The only neutral state is Colorado, leaning toward neither party; instead, Colorado is expected to vote like the nation as a whole.

Figure 5: Unemployment Rate in Percent by State in 2020 (Source: Own Representation)

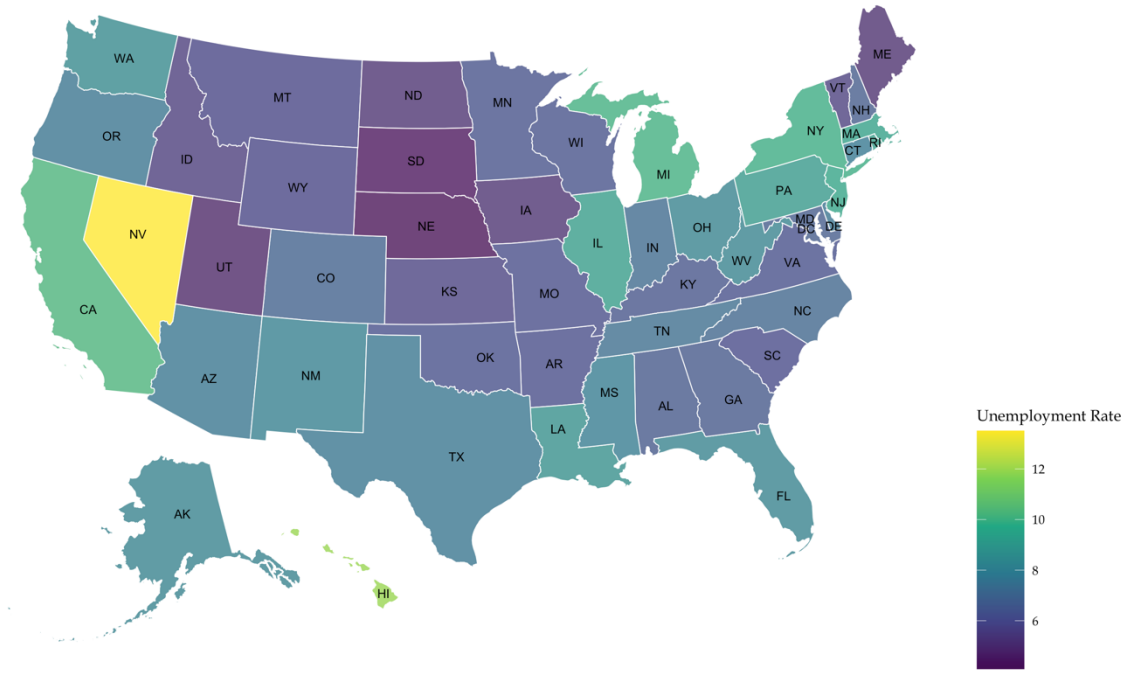
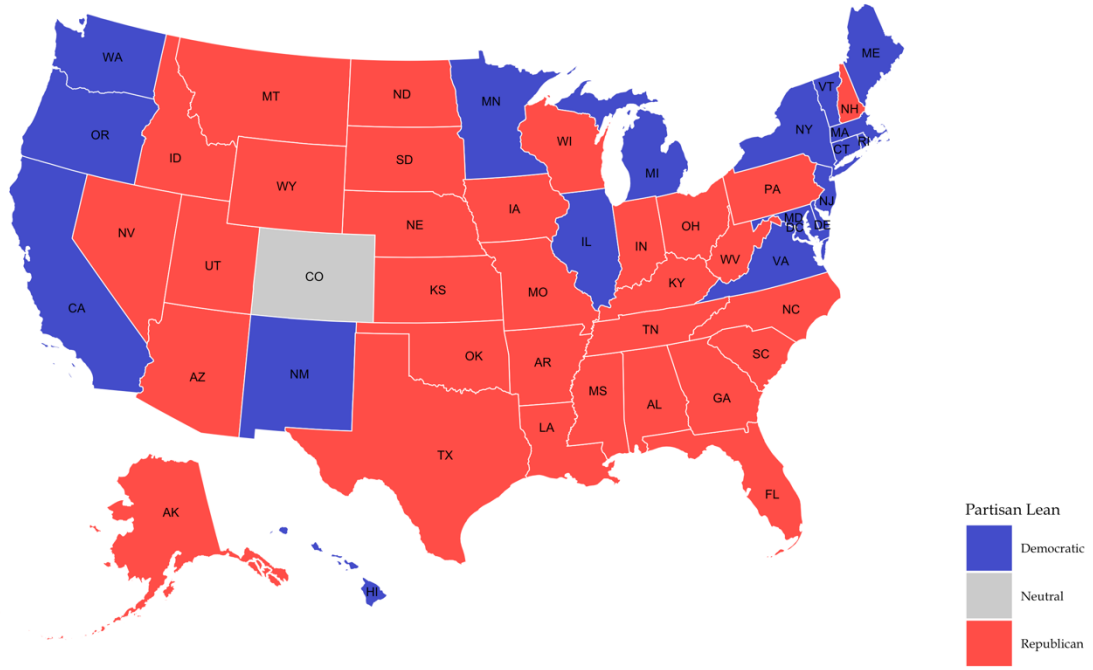


Figure 6: Partisan Lean by State in 2020 (Source: Own Representation)



After the univariate inspection, the bivariate relationships between the independent variables and the outcome are analyzed. Table 3 shows bivariate statistics for the individual-level determinants and candidate preference.[5] The mean difference is displayed for the continuous variables, while Cramer's V is used to assess the relationship between the categorical predictors and the binary outcome. The mean differences of white identity and national identity for people preferring Donald Trump or Joe Biden are positive and statistically significant at the 5% level. The mean difference of 0.29 in white identity indicates that the importance of being white is, on average, 0.29 points higher in the group of people preferring Donald Trump. The relationship is even stronger regarding the importance of being American to one's identity. On average, the importance of being American is 0.96 points higher among those preferring Donald Trump. Like the mean differences for continuous variables, the Cramer's V values for the categorical determinants are all statistically significant at the 5% level. The strongest relationship exists between the perception of the national economy and the candidate preference with a value of 0.52. This corresponds to a strong relationship,

[5] Descriptive bar plots for the continuous individual level determinants grouped by candidate preference can be found in Appendix 5.

according to Cohen (1988). However, the relationship between religious identity and candidate preference can be considered medium with a Cramer's V of 0.39.

Table 3: Bivariate Statistics for Individual Level Determinants and the Dependent Variable "Candidate Preference"

CONTINUOUS VARIABLES		**MEAN (Preference Trump)**	**MEAN (Preference Biden)**	**MEAN DIFFERENCE**
	White Identity	*2.45*	*2.16*	*0.29**
	National Identity	*4.21*	*3.25*	*0.96**
CATEGORICAL VARIABLES		**CRAMER'S V**		
	Religious Identity	*0.39**		
	National Economy	*0.52**		
	Gender	*0.05**		
	Education	*0.18**		

Notes: Mean difference is displayed for continuous predictors. Cramer's V is displayed for categorical predictors. '' indicates a p-value of < 0.05. n = 4474.*

Since Cramer's V is a measure of association between categorical variables ranging from 0 to 1, it does not tell if a relationship is positive or negative. Additionally, the single value on its own cannot tell where the relationship comes from. While one can conclude from Table 3 that there are medium to strong relationships between religious identity or the perception of the national economy and candidate preference which is even statistically significant, there is no indication of what the relationship really looks like. Figures 7 and 8 show bar plots of the categorical predictors grouped by candidate preference to address this problem and get further insights. Two conclusions can be made from the grouped frequencies of religious identities in Figure 7. First, the relationship seems to come from the strong performance of Donald Trump among Evangelicals and other Christians. Second, among the groups of non-religious people and people identifying with another religion, most respondents prefer Joe Biden. Moreover, Figure 8 sheds light on the relationship between the perception of the national economy and candidate preference. Almost everybody who feels like the national economy has improved over the past year prefers Donald Trump. In contrast, around three-quarters of respondents assessing the national economy has gotten worse prefer Joe Biden.

Figure 7: Bar Plot for Religious Identity Grouped by Candidate Preference

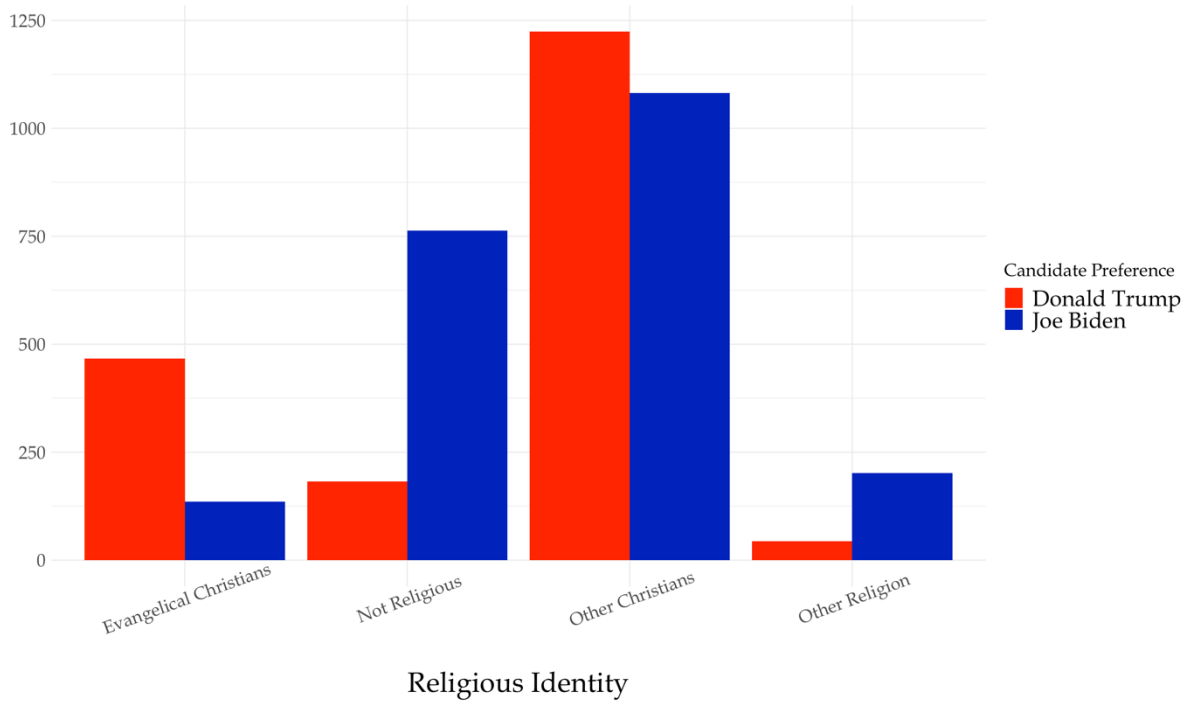
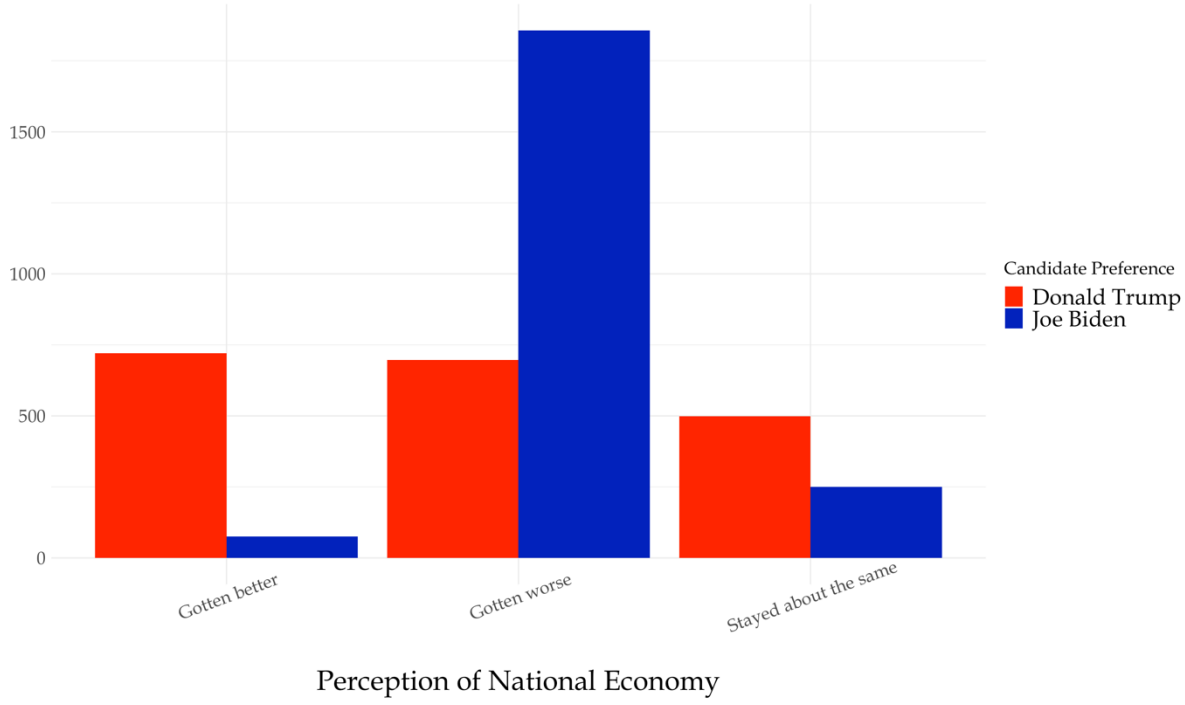


Figure 8: Bar Plot for Perception of the National Economy Grouped by Candidate Preference



Shifting to the contextual level, Table 4 shows the bivariate relationships between the contextual level predictors and candidate preference.[6] Since the original outcome is at the individual level, the candidate preference has been aggregated to the state level for better analysis. The bivariate relationships are calculated between the three predictors – ethnic diversity, unemployment, and partisan lean – and the share of people preferring Joe Biden and Donald Trump. Since the sum of shares of Joe Biden and Donald Trump equals 100% in every state, the values for Pearson's r and the mean difference are inverse values. Pearson's r has been calculated for the two continuous variables and the share of the preferred candidate in a state. Ethnic diversity negatively affects the share of people preferring Joe Biden, while the unemployment rate has a positive association. However, both effects fail to be statistically significant at the 5% level. The positive mean difference indicates a statistically significant higher share of people preferring Joe Biden in states leaning toward the Democratic Party. The inverse relationships are valid for the share of people preferring Donald Trump in a state.

Table 4: Bivariate Statistics for Contextual Level Determinants and Aggregated Candidate Preference as Share of People Preferring Joe Biden and Donald Trump in a State

DEPENDENT VARIABLE: Share of Biden Preference in a State		**PEARSON'S R**		
	Ethnic Diversity	*-0.05*		
	Unemployment Rate	*0.23*		
		MEAN (Democratic States)	**MEAN (Republican States)**	**MEAN DIFFERENCE**
	Partisan Lean	*0.67*	*0.42*	*0.25**

DEPENDENT VARIABLE: Share of Trump Preference in a State		**PEARSON'S R**		
	Ethnic Diversity	*0.05*		
	Unemployment Rate	*-0.23*		
		MEAN (Democratic States)	**MEAN (Republican States)**	**MEAN DIFFERENCE**
	Partisan Lean	*0.33*	*0.58*	*-0.25**

Notes: '' indicates a p-value of < 0.05. n = 50. Partisan lean only has n = 49 because the state of Colorado was omitted for the mean difference between Democratic-leaning and Republican-leaning states since it does not lean in favor of one party.*

[6] Further visualizations on the bivariate relationships at the contextual level with scatterplots and jittered boxplots can be found in Appendix 5.

Looking at the bivariate relationships between all independent variables and the dependent variable gives the first indication of which hypotheses might hold during the following multilevel analysis and which hypothesis might get rejected. Starting with the four hypotheses at the individual level, all hypotheses find support in the bivariate analyses. The means of white identity and national identity are higher among those preferring Donald Trump, pointing in the same direction as the hypotheses. Moreover, both mean differences are statistically significant. The data seems to support the hypotheses that people with stronger white and national identities are more likely to prefer Donald Trump over Joe Biden. Additionally, the bivariate analysis has shown that Donald Trump performs best among people identifying as Evangelicals and other Christians and those perceiving the national economy has improved over the past year. In contrast, Joe Biden performed best among non-religious people and those identifying with another religion, as well as people perceiving the national economy has worsened over the past year. Again, this points in the direction of the hypotheses. Evangelicals and other Christians in the sample have a higher likelihood of preferring Donald Trump, while people feeling like the national economy has gotten worse compared to the previous year have a higher likelihood of preferring Joe Biden. At the contextual level, the bivariate analyses provide mixed results. The correlation coefficients for ethnic diversity and unemployment rate are not statistically significant. And while the correlation between the unemployment rate and the preference shares points in the same direction as the hypothesis, the correlation between ethnic diversity and preference shares points in the opposite direction of the hypothesis. Only the mean difference of preference shares between Democratic-leaning and Republican-leaning states are statistically significant and as the hypothesis expects, people living in a Democratic-leaning state seem to have an overall higher likelihood of preferring Joe Biden than those in a Republican-leaning state. Nonetheless, this can only be a first look at the hypotheses. The evaluation of hypotheses and consequences for the theoretical model and research questions are based on the multilevel logistic regression in the following chapter.

5.2. Multilevel Analysis

The results of the multilevel logistic regression are displayed in Table 5. It includes the logit coefficients and standard errors for the independent variables. Since logit coefficients are hardly interpretable, predicted probabilities and average marginal effects will be provided in the second step of the analysis. Four models are shown in Table 5. First is the Null Model with only the intercept. Second is the Identity Model, including the three identity variables: white identity, national identity, and religious identity. Third is the Individual Model, including the

identity model as well as the perception of the national economy and control variables. Fourth is the Full Model, including the Individual Model and the contextual level predictors: ethnic diversity, unemployment rate, and partisan lean. The dependent variable in all models is the binary candidate preference between Donald Trump (0) and Joe Biden (1). Moreover, Table 5 includes the variance component of the random effects and intra-class correlation. Marginal R^2, conditional R^2, and BIC are displayed for model comparison.

The Identity Model includes standardized white identity and national identity as well as a categorical religious identity variable with the reference category being "Not Religious". The positive coefficient for white identity indicates a stronger white identity being associated with a higher likelihood of preferring Joe Biden, ceteris paribus. An increase in white identity by one standard deviation leads to an increase in the predicted logit of 0.13 points. The effect of white identity is also highly statistically significant at the 0.1% level. The likelihood of finding this or an even more extreme result, although the null hypothesis of a coefficient = 0 is true for the population, is below 0.1%. This result is contrary to the insights gained from the bivariate analysis, which indicated that a stronger white identity is associated with a higher likelihood of preferring Donald Trump. It also points in the opposite direction of H_1. This contradicting result will be further analyzed later. The national identity coefficient of -0.77 shows a stronger national identity being associated with a lower likelihood of preferring Joe Biden and, in converse conclusion, a higher likelihood of preferring Donald Trump. Again, the coefficients are statistically significant at the 0.1% level. The result for national identity in the model supports H_2. The coefficients for religious identity must be interpreted in reference to the category "Not Religious". Therefore, people identifying as Evangelical (-2.18) or other Christians (-1.21) are more likely to prefer Donald Trump than those who are not religious. The effect of identifying as Evangelical is even stronger. Both effects are statistically significant. Moreover, people identifying with another religion are more likely to prefer Joe Biden, although the effect is not statistically significant. The results from the Identity Model support H_3 that Evangelicals and other Christians have a higher likelihood of preferring Donald Trump than the other groups.

The Individual Model extends the Identity Model by the perception of the national economy and the control variables education and gender. The coefficients for the perception of the national economy must be interpreted in reference to the category "Stayed the same". On the one hand, the negative coefficient for people who perceived the national economy has improved over the past year indicates these people have a higher likelihood of preferring Donald Trump than those who feel the national economy stayed about the same. On the other hand, people who feel the national economy has worsened are more likely to prefer Joe Biden.

Table 5: Regression Table for the Dependent Variable: Favoring Donald Trump (0) or Joe Biden (1)

	Null Model		**Identity Model**		**Individual Model**		**Full Model**	
	Logit	*Std. Error*	*Logit*	*Std. Error*	*Logit*	*Std. Error*	*Logit*	*Std. Error*
Intercept	0.05	0.09	1.14 ***	0.11	-0.13	0.15	-0.37 **	0.14
Individual Level								
White Identity			0.13 ***	0.04	0.14 ***	0.04	0.14 **	0.04
National Identity			-0.77 ***	0.04	-0.71 ***	0.05	-0.70 ***	0.05
Religious Identity: Ref. Not Religious								
Evangelical			-2.18 ***	0.13	-2.06 ***	0.15	-2.04 ***	0.15
Other Christian			-1.21 ***	0.10	-1.08 ***	0.11	-1.07 ***	0.11
Other Religion			0.15	0.19	0.14	0.22	0.14	0.22
National Economy: Ref. Stayed the same								
Gotten Better					-1.44 ***	0.15	-1.43 ***	0.15
Gotten Worse					1.54 ***	0.10	1.53 ***	0.10
State Level								
Ethnic Diversity							-0.13 **	0.05
Unemployment Rate							0.02	0.05
Partisan Lean: Ref. Republican								
Democratic							0.64 ***	0.11
Neutral							0.17	0.27
Control Variables	No		No		Yes		Yes	
Random Effects								
τ_{00}	0.32 $_{state}$		0.16 $_{state}$		0.12 $_{state}$		0.01 $_{state}$	
ICC	0.09		0.05		0.04		0.00	

N	50_{state}	50_{state}	50_{state}	50_{state}
Observations	4474	4474	4474	4474
Marginal R^2 / Conditional R^2	0.000 / 0.089	0.286 / 0.318	0.514 / 0.531	0.541 / 0.542
BIC	6005.0	5061.2	4153.2	4152.6

*Notes: All continuous variables in the model were standardized with mean = 0 and standard deviation = 1. The control variables included are gender and education. Both variables are coded categorically. Gender consists of the categories "Female" and "Male" with "Male" being the reference category. Education consists of the categories "College Degree" and "No College Degree" with "No College Degree" being the reference category. Therefore, the intercept is interpreted as log odds for candidate preference if all continuous variables are at their mean and all categorical variables are at their reference category. * $p<0.05$ ** $p<0.01$ *** $p<0.001$.*

Both effects are statistically significant and support H_4 that people evaluating the economy as worse are more likely to prefer Joe Biden over Donald Trump. Looking at the identity variables, the inclusion of the perception of the national economy and the control variables only resulted in slight changes in the estimates and no changes in significance. Therefore, the effects of social identities still hold if one of the strongest and most important predictors for political attitudes is included.

The Full Model includes all variables at the individual level as well as ethnic diversity, unemployment rate, and partisan lean at the contextual level. The estimate for ethnic diversity is negative and statistically significant at the 1% level. Hence, the model indicates that living in a more ethnically diverse state leads to a lower likelihood of preferring Joe Biden. H_5, based on the contact hypothesis, stated that living in a more diverse state should lead to a higher likelihood of preferring Joe Biden and must be rejected. H_6 hypothesizes a positive relationship between the unemployment rate and candidate preference. The coefficient for the unemployment rate of 0.02 points in the direction of the hypothesis but is very small and statistically insignificant. Thus, H_6 must be rejected as well. The coefficients for partisan lean indicate that people living in a Democratic-leaning state have a higher likelihood of preferring Joe Biden than those living in a Republican-leaning state. This effect is statistically significant and supports H_7. The Neutral-leaning state's result is statistically insignificant since only one state was not leaning toward either party in 2020. There are only very slight changes to the individual level coefficients and p-values compared to the Individual Model. All interpretations of effects from previous models remain valid in the Full Model.

Additionally, Table 5 provides information on the random effects at the state level. These indicate how much the outcome differs in each of the 50 states. τ_{00} of 0.32 means that the 50 states differ the most in the Null Model without any predictor variables. This difference

decreases if individual-level predictors are included in the Identity Model and Individual Model to 0.16 and 0.12, respectively. However, the differences between the states almost vanish to 0.01 when the state-level predictors are included in the Full Model. The intra-class correlation paints a similar picture, with around 9% of the variance in the outcome being attributed to the state level in the Null Model. While the individual-level models already decrease this value to about 5% and 4%, the ICC does not reach a value of around 0 until the contextual-level predictors are included. This reduction in the ICC can be attributed primarily to ethnic diversity and partisan lean because of the statistically significant effects found in multilevel logistic regression.

Indicators for model comparison can be found at the bottom of Table 5. The marginal R^2 and conditional R^2 suggested by Nakagawa and Schielzeth to assess explained variance in generalized linear mixed models and the BIC are displayed in the table. The marginal R^2 indicates how much variance is explained by the fixed effects. In contrast, the conditional R^2 is an indicator for how much variance is explained by the whole model – fixed and random effects (Nakagawa and Schielzeth 2013: 137). Both indicators range from 0 to 1. Since the Null Model does not contain any fixed effects, the marginal R^2 is 0. The conditional R^2 in the Null Model is equal to the ICC because the random intercepts for states are the only component in the model yet. Both values increase as the models become more complex, with the Full Model explaining most of the variance in the outcome. This speaks for the Full Model as the relative best model in this comparison. However, there are only slight improvements from the Individual Model to the Full Model. The BIC also sees a slight improvement from 4153.2 in the Individual Model to 4152.6 in the Full Model. An additional likelihood ratio test between the Individual Model and the Full Model produced a statistically significant result.[7] Meaning that the null hypothesis of the three contextual-level effects simultaneously being 0 can be rejected, and the Full Model eventually stands as the relative best model. The almost identical marginal R^2 and conditional R^2 in the Full Model indicate that nearly all explained variance is due to the fixed effects. There seems only to be a small contribution left by the random intercept varying across the 50 states. This is in line with the development of the ICC over the models. Almost no variance in candidate preference between the states is left if ethnic diversity, unemployment rate, and partisan lean are included as contextual-level predictors.

To check for the robustness of the result, two random effects models with all predictors from the Full Model and the original feeling thermometer variables as dependent variables have

[7] The results of the likelihood ratio test can be found in Appendix 6.

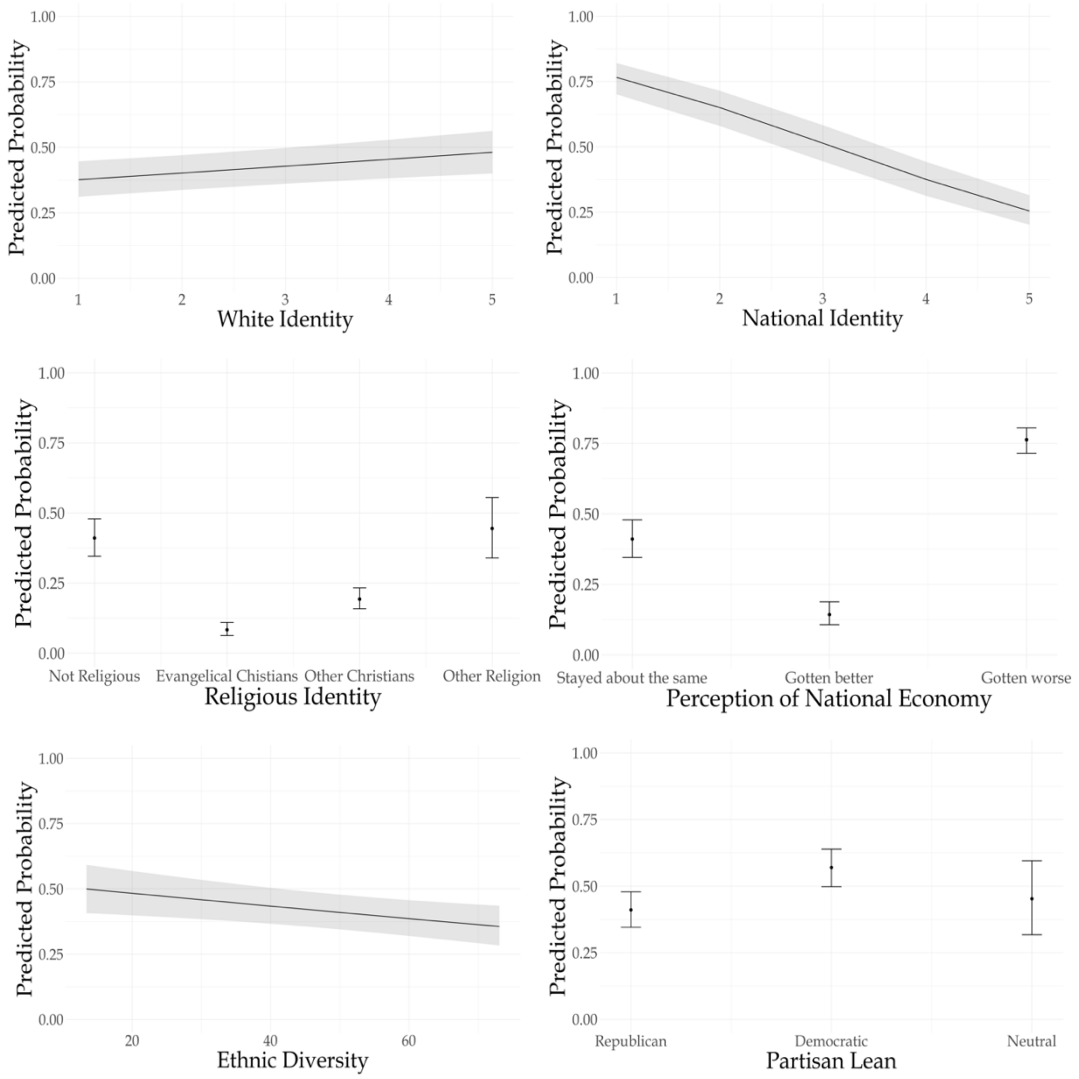
been calculated.[8] Since preferring Joe Biden was coded as 1, and Donald Trump was coded as 0, the effects in the model with the feeling thermometer towards Biden as the outcome should show the same directions as the Full Model in Table 5. The model with the feeling thermometer toward Donald Trump should show opposite effect directions. The robustness checks delivered the expected results with no changes in direction or statistical significance. This strengthens the results in Table 5 because the results are not biased due to the transformation of the outcome variable.

Predicted probabilities for all effects that proved statistically significant in the Full Model are displayed in Figure 9 to get further insights into the effect sizes of the independent variables. They are better suited to interpret the effect sizes than logit coefficients because they result in more intuitively understandable interpretations. Predicted probabilities allow for statements on how changes in the independent variables affect the probability of the outcome being 1. The effects of white identity, national identity, religious identity, perception of the national economy, ethnic diversity, and partisan lean can be seen in Figure 9. The graphs show the predicted probability of preferring Joe Biden for every value of the independent variables while all other variables are held constant. Continuous variables are held constant at their mean, while categorical variables are held constant at the reference level. Predicted probabilities allow for statements on the effect sizes. Holding everything else constant, a person with a weak white identity of 1 has a predicted probability of around 38%. In comparison, a person with a strong white identity of 5 has a predicted probability of about 48%. This corresponds to a difference of ten percentage points between the weakest and strongest white identities. In contrast, the difference between the weakest and strongest national identities is about 50 percentage points. A person to whom it is very important to be American has a predicted probability of preferring Joe Biden of around 26%, while somebody to whom it is not at all important has a predicted probability of about 76%. This shows a difference in effect size between white identity and national identity of about 40 percentage points, with the national identity effect being stronger. The predicted probabilities for religious identity show that Evangelicals (around 8%) and other Christians (around 19%) have the lowest predicted probability, and people who identify with another religion or as not religious have higher values of around 45% and 41%, respectively. The 95% confidence interval shows a statistically significant difference between Evangelicals, other Christians, and the other categories. However, the predicted probabilities of not religious people and those identifying with another religion do not differ statistically significantly. In contrast, the probabilities for all three categories of perception of the national economy show statistically significant

[8] The regression tables for the robustness checks can be found in Appendix 7.

differences. People with the perception that the economy has gotten better over the past year have the lowest probability of preferring Joe Biden at around 10%, while people feeling like it stayed about the same are at around 41%, and people feeling like it has gotten worse are at around 76%. At the contextual level, there is a difference in predicted probabilities of around 15 percentage points between the state with the lowest and the state with the highest ethnic diversity. The predicted probability of preferring Joe Biden in the state with the lowest ethnic diversity is around 50%, and in the state with the highest ethnic diversity is only around 35%. The graph shows a statistically significant difference between Republican-leaning and Democratic-leaning states for partisan lean. As expected, the predicted probabilities show that it is more likely to prefer Donald Trump in a Republican-leaning state and Joe Biden in a Democratic-leaning state since the predicted probability in Republican-leaning states (41%) is below and in Democratic-leaning states (54%) is above 50%.

Figure 9: Predicted Probabilities for All Statistically Significant Effects in the Multilevel Analysis



Notes: White Identity and National Identity are measured on a scale from 1 to 5. Ranging from 1 "Not at all important" to 5 "Extremely important". Ethnic Diversity is measured from 0 to 100. 95%confidence intervals are depicted in all figures.

In the last step, average marginal effects for the Identity Model, Individual Model, and Full Model are provided in Table 6. Although average marginal effects are related to predicted probabilities, they allow for comparison of effects across models. Thus, average marginal effects are used to get further insights into how the effects change between the different models. The average marginal effect of white identity in the Identity Model indicates that, on average, the predicted probability of preferring Joe Biden increases by two percentage points if white identity increases by one unit. The effect of white identity is stable over all models. The effect of national identity is strongest in the Identity Model with 0.12 and loses some strength when the perception of the national economy and the control variables are included. In the Full Model, the predicted probability of preferring Joe Biden decreases by eight percentage points on average if national identity is increased by one unit. A similar pattern can be observed for religious identity. While the effect sizes drop from the identity to the Individual Model, they remain nearly the same between the Individual Model and the Full Model. On average, the predicted probability of preferring Joe Biden is 33 percentage points lower for Evangelicals than for not religious people in the Full Model. In comparison, it is 44 percentage points lower in the Identity Model. The effect for other Christians is weaker, with a difference of 16 percentage points compared to not religious people in the Full Model and 25 percentage points in the Identity Model. The effect size of the perception of the national economy stays nearly the same between the models without and with contextual-level predictors. People perceiving the national economy as better than a year ago, on average, have a predicted probability of 22 percentage points lower than those who felt like it stayed the same. The predicted probability of preferring Joe Biden is 28 percentage points higher for those who think the national economy has gotten worse in the Full Model. Looking at the contextual level, the average marginal effects of ethnic diversity and the "Democratic" category of the partisan lean variable show statistically significant results. On average, the predicted probability of preferring Joe Biden decreases by 0.14 percentage points if the ethnic diversity in a state is increased by one point. And compared to a Republican-leaning state, a Democratic-leaning state has a predicted probability of preferring Joe Biden which is 10 percentage points higher on average.

The most striking result of the multilevel analysis is the positive effect of white identity on candidate preference in the model. An increase in the strength of white identity has proven to be associated with a higher likelihood of preferring Joe Biden instead of Donald Trump. This leads to rejecting the hypothesis based on Social Identity Theory and Donald Trump's efforts to raise the salience of white identity during the 2020 election campaign. Additionally, this result contradicts previous research on support for Donald Trump in the 2016 U.S. Presidential

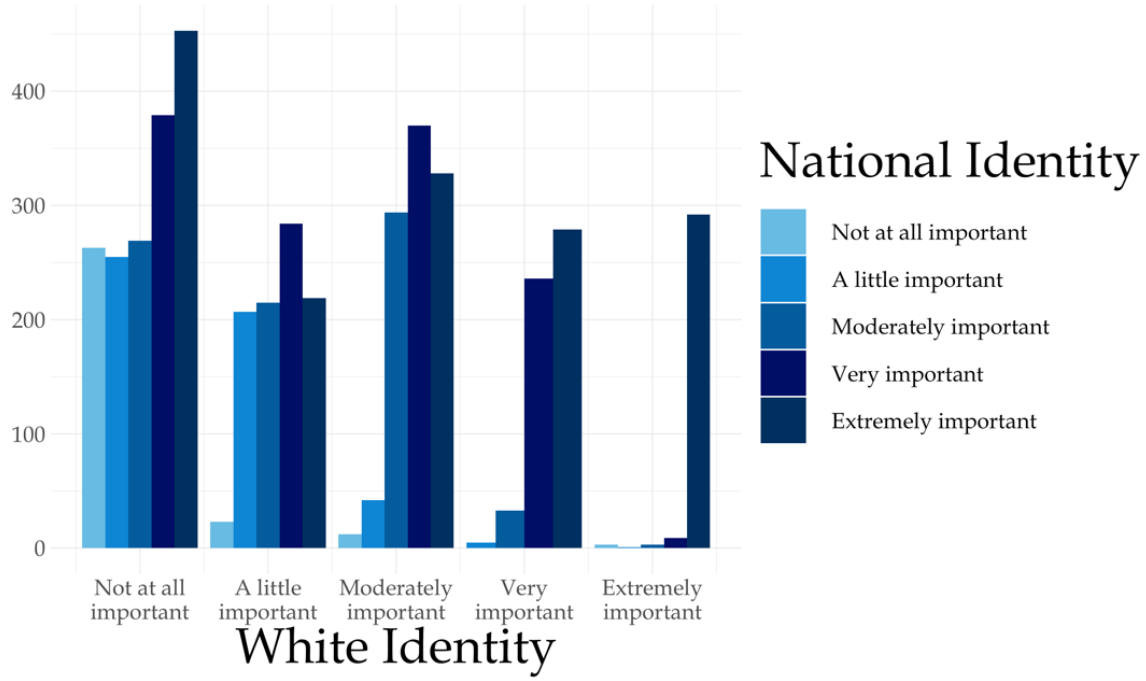
Table 6: Average Marginal Effects (AME) for All Variables Included in the Multilevel Analysis

	Identity Model	**Individual Model**	**Full Model**
	AME *(Std. Error)*	*AME* *(Std. Error)*	*AME* *(Std. Error)*
Individual Level			
White Identity	0.02*** (0.01)	0.02*** (0.00)	0.02*** (0.00)
National Identity	-0.12*** (0.01)	-0.08*** (0.01)	-0.08*** (0.01)
Religious Identity: Ref. Not Religious			
Evangelical	-0.44*** (0.02)	-0.33*** (0.02)	-0.33*** (0.02)
Other Christian	-0.25*** (0.02)	-0.17*** (0.02)	-0.16*** (0.02)
Other Religion	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
National Economy: Ref. Stayed the same			
Gotten Better		-0.22*** (0.02)	-0.22*** (0.02)
Gotten Worse		0.29*** (0.02)	0.28*** (0.02)
State Level			
Ethnic Diversity			-0.00*** (0.00)
Unemployment Rate			0.00 (0.00)
Partisan Lean: Ref. Republican			
Democratic			0.10*** (0.02)
Neutral			0.03 (0.04)

*Notes: Results are based on models without scaled continuous variables. Control variables are included for the Individual Model and the Full Model. Standard errors in parentheses. * p<0.05 ** p<0.01 *** p<0.001.*

Election (Sides, Tesler, and Vavreck 2017: 3). However, this is not an entirely novel result. Buyuker and colleagues found no effect or even a negative effect of white identity on vote choice for Donald Trump in relation to Hillary Clinton in 2016 (Buyuker et al. 2021: 602). This suggests further research on how white identity works in the contest between two white candidates who draw attention to different ethnic identities during their campaign.

Figure 10: Bar Plot of White Identity Grouped by National Identity



But because the bivariate relationship between white identity and candidate preference pointed in the expected direction and the effect changed directions in the regression framework, the analysis allows us to see which identity in the model caused the change. Only if national identity is included beside white identity does the white identity effect change its direction. However, including an interaction effect of white identity and national identity did not result in a statistically significant relationship.[9] Nevertheless, Figure 10 shows some interesting descriptive relationships between the two variables. Almost everybody in the sample who holds a strong white identity also holds a strong national identity. However, almost everybody holding a weak national identity also holds a weak white identity. A Pearson's correlation coefficient test also gave a statistically significant result of 0.39 - equivalent to a medium relationship.[10] These results show that further research on the

[9] An extended multilevel regression model with an interaction effect between white identity and national identity can be found in Appendix 8.
[10] A table reporting Pearson's r on the relationship between white identity and national identity can be found in Appendix 8.

relationship between white identity and national identity in the context of electoral politics is needed. The next chapter will put the results into the context of the existing literature. Moreover, it will point out various limitations to the analysis and how to solve the problems in future research.

5.3. Discussion and Limitations

The literature review at the beginning identified two essential parts missing from the research on candidate preference in the United States. First is the missing focus of social identities. Second is the missing focus on effects at the contextual level of states. The analysis closed these research gaps by building a multilevel model including social identities at the individual level and adding ethnic diversity, unemployment rate, and partisan lean at the state level.

Hypotheses H_2, H_3, H_4, and H_7 found support in the multilevel analysis. On average, a stronger national identity, identifying as Evangelical or other Christian, and perceiving the national economy as better than a year ago increased the likelihood of preferring Donald Trump over Joe Biden and vice versa. Moreover, living in a Democratic-leaning state increased the likelihood of preferring Joe Biden on average. However, H_1, H_5, and H_6 had to be rejected. The analysis found no relationship between the unemployment rate at the state level and candidate preference. Additionally, effects opposite to the hypothesized effects were found for white identity and ethnic diversity. The white identity effect was particularly striking because preliminary bivariate analyses pointed toward the relationship formulated in H_1. But after national identity was included in the regression model, the effect reversed. The expected relationship between ethnic diversity and candidate preference was based on the contact hypothesis. Higher ethnic diversity in a state should lead to less prejudice toward outgroups since white people have more contact with these outgroups. As a result, they should favor the liberal candidate embracing other ethnic identities in his election campaign. However, the contradicting result might be an indicator of the effect of perceived status threat. Higher ethnic diversity and more contact with outgroups might lead to a perceived threat to the status of white people which may then turn them to the more conservative candidate who is perceived to protect their social status. The effect of perceived status threats is already known in the context of policy attitudes and has been tested in experimental settings with white people perceiving a threat to their status due to ethnic diversity leaning more toward conservative policies (Craig and Richeson 2014: 1195). Thus, the influence of ethnic diversity on candidate preference should be topic to further research.

The main contribution to the research on candidate preference is the consideration of the contextual level. While most research focuses on individual level aspects in explaining candidate preference in the U.S. (see e.g. Luttig 2021; King 2022; Kam 2007), the thesis combines individual and contextual level predictors. The analysis shows that (a) there is about 9% of the variance in candidate preference that can be associated with the state-level that has to be considered when looking at the candidate preference in the 2020 election, and (b) this variance can be almost completely explained when including ethnic diversity, unemployment, and partisan lean by states. In addition, the results contribute to understanding social identities' effects on candidate preference even in a context where economic considerations remain a strong predictor. Social identities prove to be important not only in situations where candidates and voters share a common identity (see Jackson 2011; Stokes-Brown 2006) but also where candidates activate certain identities.

The results also have multiple implications for science, politics, and society. First, the state level has shown to be important when it comes to candidate preference. This should motivate scholars to focus on effects from different hierarchical levels. Additionally, political decision-makers in state legislatures can influence elections by changing certain conditions in a state. Second, although some effects are small, like the effect of ethnic diversity, they remain important. Especially in the U.S. Presidential Election, an election in which a few thousand votes in a state can determine the outcome for a whole nation. Third, the white identity effect does not work as expected based on previous research (Sides, Tesler, and Vavreck 2017: 3) and the theoretical background of the saliency of social identities in the Presidential Election 2020. Most likely due to the correlation with national identity. Lubbers and Coenders (2017: 115) already found a relationship between national identity and ethnic threat in support for radical right-wing parties in Europe. This should lead to more research in that area in the U.S. and a shift in society when thinking about the partisan link of white identity to the Republican Party. Considering an ethnic-nationalist approach, as proposed in the literature on European radical right parties (see e.g. Bonikowski 2017), where people identify as White Americans might be a concept for further research. Fourth, social identities are important predictors of candidate preference, even when sociotropic considerations are accounted for. Thus, scholars should not neglect social identities when explaining phenomena in contemporary American politics.

However, this study comes with a variety of limitations regarding data and research design. This section focuses on constraints to the analysis based on the sampling, measurement, and availability of data, as well as limitations to the analysis method and causality. Three limitations of the analysis can be attributed to the nature of the data. First, states are still very large geographical entities with highly diverse areas. However, there are no datasets with

samples large enough to break down context factors at the county level. Therefore, the results at the state level might be biased by too much variation within each state. Second, the sample size of other ethnic identities in the ANES 2020 is too small to conduct analyses beyond white identity. The analysis could not control for varying effects across different ethnic identities. Third, the different social identities used in the analysis are on different scales. While white identity and national identity use the same intensity scale of how important being white or American is to someone's identity, religious identity is a simple categorical measure. This makes comparisons between the effects of the identities hardly feasible.

Regarding the research design, four limitations must be noted. First, the random intercept model is the simplest form of random effects modeling. Including more complex mechanisms like random slopes could get even more insights into the effects of the model. However, including random slopes was outside of the scope of this thesis. Second, the analysis covers only one year and the candidate preference in one election. This limits the statements based on the results to only being valid in this particular context. Third, the analysis could only get a glimpse of the effect of white identity on the preference for Joe Biden. Due to space and time constraints, a further analysis was not in the scope of this work and needs to be at the center of upcoming research. Fourth, the analysis fails to take all four hurdles of causality proposed by Kellstedt and Whitten (Kellstedt and Whitten 2013: 55). The first hurdle of a credible causal mechanism that connects social identities to candidate preference was taken with the concept of the saliency of different social identities. At the contextual level, the contact hypothesis and the Economic Voting Theory provided a plausible theoretical foundation. The second hurdle of ruling out the possibility that candidate preference could cause social identities is taken by the nature of both variables. While social identities are a long-term factor affecting social and political behavior over a long time, candidate preference is one attitude in one particular situation. And since the outcome is at the individual level, it cannot influence the contextual-level determinants included in the model. The analysis showed covariation between the social identities as well as contextual-level predictors and candidate preference which checks the box of the third hurdle. However, since the analysis is conducted with observational data, one cannot be sure to control for every possible confounding variable. This limits the analysis' ability to take the fourth hurdle of controlling for every confounder. Some control variables were included, but the drawback of every non-experimental study is omitted variable bias.

These limitations can be addressed in future research using different data types. Using panel data or even experimental data would address the problem of causality and deliver more robust results regarding the causal effects. Datasets with larger subsamples of other ethnic groups would allow for more diverse analyses of ethnic identities. Using datasets with a

smaller geographical scope than the whole U.S. would allow for more detailed data even at the county level to have more homogeneous contexts. Furthermore, a more complex analysis, including random slopes or cross-level interactions, might shed new light on the effects found in the random intercept model. In addition, future research should focus on theoretical and empirical work on the white identity effects on candidate preference to disentangle the underlying mechanisms.

6. Conclusion

Previous research has missed out on the relationships between social identities and candidate preference in U.S. Presidential Elections. Moreover, a systematic analysis of state-level determinants for candidate preference in U.S. Presidential Elections was missing in Political Science research. But what does the analysis of the U.S. Presidential Election 2020 reveal about the relationships between social identities and candidate preference when factoring in the state level?

Individual-level data from the American National Elections Study 2020 and contextual-level data from different sources were used in a multilevel logistic regression model with a random intercept. At the individual level, the focus was on respondents' white identity, national identity, and religious identity. National identity and religious identity provided the expected effects. The more important being American was to someone's identity, the more likely they were to prefer Donald Trump over Joe Biden. People identifying as Christians had a statistically significantly higher likelihood of preferring Donald Trump over Joe Biden than other groups, with Evangelicals having the highest likelihood. However, white identity had a different effect from what was expected based on the theoretical model and previous research. The more important being white was to someone's identity, the more likely they were to prefer Joe Biden over Donald Trump. A relationship between white and national identity was found but could not enhance the statistical models. These results show that further research on the relationship between white and national identity in electoral politics is needed. Moreover, all social identity effects on candidate preference hold if the models include one of the most important predictors for political attitudes and behavior: sociotropic considerations.

At the contextual level, the analysis looked at ethnic diversity, unemployment rate, and partisan lean in each of the 50 states. Higher ethnic diversity in a state statistically significantly decreased the likelihood of preferring Joe Biden over Donald Trump, while the unemployment rate did not affect the candidate preference. However, the effect of partisan lean was as expected. People living in a state leaning more toward the Democratic Party than the whole country had a higher likelihood of preferring Joe Biden over Donald Trump than those living in a Republican-leaning state. Furthermore, the multilevel analysis revealed that almost no variance in candidate preference between the states is left when ethnic diversity, unemployment rate, and partisan lean are included in the model. The intra-class correlation coefficient detected a variance in candidate preference between the 50 states of about 9% without any predictors.

Three main findings must be highlighted regarding the research questions posed in the introduction. First, social identities influence candidate preference between Joe Biden and Donald Trump. White identity, national identity, and religious identity statistically significantly affected which candidate people prefer. A stronger white identity is associated with an increased likelihood of preferring Joe Biden. In comparison, a stronger national identity and a Christian identity are associated with an increased likelihood of preferring Donald Trump. Second, the state level proved important in explaining candidate preference in the U.S Presidential Election 2020. Around 9% of the variance in candidate preference could be attributed to the state level and was only completely resolved when contextual-level predictors were included. Third, ethnic diversity and partisan lean at the state level showed statistically significant effects on candidate preference. Lower ethnic diversity and living in a Democratic-leaning state were associated with an increased likelihood of preferring Joe Biden. Higher ethnic diversity and living in a Republican-leaning state were associated with an increased likelihood of preferring Donald Trump.

However, there are limitations to the answers to the research questions. States are still very large geographical entities with highly diverse regions, which might skew the results at the contextual level. Perhaps, more detailed analyses on a lower level would come to different conclusions regarding contextual-level predictors. Additionally, not all ethnic identities of interest were part of the analysis. Due to small subsamples of ethnic groups other than people identifying as white, the analysis was limited to white identifiers. However, the effects of national or religious identity might differ between different ethnic groups. The most important limitation is the limitation to causality. All statements on the effects of social identities at the individual level and other predictors at the contextual level are based on correlations. The observational data used does not allow for conclusions regarding causality.

Therefore, this work motivates further research projects. Collecting data with larger subsamples of certain ethnic groups would allow for more diverse analyses of the effects of ethnic identities. It might even disentangle the underlying mechanisms of the white identity effect. And experimental research would allow for more robust statements about the causal impact between predictors and candidate preference. A time-series analysis would allow for statements over time and how the effects of individual-level and contextual-level predictors of candidate preference might change. And in the end, hypotheses regarding social identities and candidate preference, as well as the influence of contextual levels, can also be applied to contexts other than the United States.

References

Abbe, Owen G., Jay Goodliffe, Paul S. Herrnson, and Kelly D. Patterson. 2003. 'Agenda Setting in Congressional Elections: The Impact of Issues and Campaigns on Voting Behavior', *Political Research Quarterly*, 56: 419-30.

Adida, Claire, Jessica Gottlieb, Eric Kramon, and Gwyneth McClendon. 2017. 'Reducing or Reinforcing In-Group Preferences? An Experiment on Information and Ethnic Voting', *Quarterly Journal of Political Science*, 12: 437-77.

Ahler, Douglas J. 2018. 'The Group Theory of Parties: Identity Politics, Party Stereotypes, and Polarization in the 21st Century', *The Forum*, 16: 3-22.

Allport, Gordon W. 1954. *The nature of prejudice* (Addison-Wesley Publishing Company: Cambridge, Mass.).

American National Election Studies. 2021. "ANES 2020 Time Series Study Full Release [dataset and documentation]. July 19, 2021 version." In.

Andersen, Robert, and Anthony Heath. 2003. 'Social identities and political cleavages: the role of political context', *J Royal Statistical Society*, 166: 301-27.

André, Audrey, and Sam Depauw. 2017. 'Too much choice, too little impact: a multilevel analysis of the contextual determinants of preference voting', *West European Politics*, 40: 598-620.

Bartels, Larry M. 2002. 'Beyond the Running Tally: Partisan Bias in Political Perceptions', *Political Behavior*, 24: 117-50.

Ben-Bassat, Avi, and Momi Dahan. 2012. 'Social Identity and Voting Behavior', *Public Choice*, 151: 193-214.

Berliant, Marcus, and Hideo Konishi. 2005. 'Salience: Agenda choices by competing candidates', *Public Choice*, 125: 129-49.

Bienenstock, Elisa Jayne, Phillip Bonacich, and Melvin Oliver. 1990. 'The effect of network density and homogeneity on attitude polarization', *Social Networks*, 12: 153-72.

Bonikowski, Bart. 2017. 'Ethno-nationalist populism and the mobilization of collective resentment', *The British Journal of Sociology*, 68: S181-S213.

Boswijk, Vincent, and Matt Coler. 2020. 'What is Salience?', *Open Linguistics*, 6: 713-22.

Bradberry, Leigh A. 2016. 'The Effect of Religion on Candidate Preference in the 2008 and 2012 Republican Presidential Primaries', *PLoS ONE*, 11: 1-24.

Buyuker, Beyza, Amanda Jadidi D'Urso, Alexandra Filindra, and Noah J. Kaplan. 2021. 'Race politics research and the American presidency: thinking about white attitudes, identities and vote choice in the Trump era and beyond', *The Journal of Race, Ethnicity, and Politics*, 6: 600-41.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter* (University of Chicago Press).

Cassese, Erin C. 2020. 'Straying from the Flock? A Look at How Americans' Gender and Religious Identities Cross-Pressure Partisanship', *Political Research Quarterly*, 73: 169-83.

Cineas, Fabiola. 2020. 'How Biden has — and hasn't — harnessed the national reckoning on race', *Vox*.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences* (Lawrence Erlbaum Associates: Hillside, NJ).

Craig, Maureen A., and Jennifer A. Richeson. 2014. 'On the Precipice of a "Majority-Minority" America: Perceived Status Threat From the Racial Demographic Shift Affects White Americans' Political Ideology', *Psychological Science*, 25: 1189-97.

Devine, Christopher J. 2015. 'Ideological Social Identity: Psychological Attachment to Ideological In-Groups as a Political Phenomenon and a Behavioral Influence', *Political Behavior*, 37: 509-35.

Elving, Ron. 2018. 'The Florida Recount Of 2000: A Nightmare That Goes On Haunting', *NPR*.

Enns, Peter K., and Paul M. Kellstedt. 2008. 'Policy Mood and Political Sophistication: Why Everybody Moves Mood', *British Journal of Political Science*, 38: 433-54.

Erikson, Robert S., Olle Folke, and James M. Snyder Jr. 2015. 'A Gubernatorial Helping Hand? How Governors Affect Presidential Elections', *The Journal of Politics*, 77: 491-504.

Falkinger, Josef. 2008. 'LIMITED ATTENTION AS A SCARCE RESOURCE IN INFORMATION-RICH ECONOMIES', *The Economic Journal*, 118: 1596–620.

Finch, W. Holmes, Jocelyn E. Bolin, and Ken Kelley. 2019. *Multilevel Modeling in R* (Taylor & Francis Group, LLC: New York).

FiveThirtyEight. 2020. 'FiveThirtyEight's Partisan Lean (2020)', Accessed 19.09.2022. https://github.com/fivethirtyeight/data/tree/master/partisan-lean/2020.

Garzia, Diego, and Frederico Ferreira da Silva. 2022. 'The Electoral Consequences of Affective Polarization? Negative Voting in the 2020 US Presidential Election', *American Politics Research*, 50: 303-11.

Giles, Michael W., and Kaenan Hertz. 1994. 'Racial Threat and Partisan Identification', *The American Political Science Review*, 88: 317-26.

Gomez, Brad T. , and J. Matthew Wilson. 2001. 'Political Sophistication and Economic Voting in the American Electorate: A Theory of Heterogeneous Attribution', *American Journal of Political Science*, 45: 899-914.

Green, Donald Philip, and Jonathan A. Cowden. 1992. 'Who Protests: Self-Interest and White Opposition to Busing', *The Journal of Politics*, 54: 471-96.

Greene, Steven. 2004. 'Social Identity Theory and Party Identification', *Social Science Quarterly*, 85: 136-53.

Hamilton, David L., and Sara A. Crump. 2004. 'Stereotypes.' in Charles D Spielberger (ed.), *Encyclopedia of Applied Psychology* (Elsevier).

Healy, Andrew J., Mikael Persson, and Erik Snowberg. 2017. 'Digging into the Pocketbook: Evidence on Economic Voting from Income Registry Data Matched to a Voter Survey', *The American Political Science Review*, 11: 771-85.

Higgins, E. Tory. 1996. 'Knowledge activation: Accessibility, applicability, and salience.' in E. Tory Higgins and A. W. Kruglanski (eds.), *Social psychology: Handbook of basic principles* (The Guilford Press).

Hoffner, Cynthia, and Raiza A. Rehkoff. 2011. 'Young Voters' Responses to the 2004 U.S. Presidential Election: Social Identity, Perceived Media Influence, and Behavioral Outcomes', *Journal of Communication*, 61: 732-57.

Hook, Janet, and Melanie Mason. 2020. 'Biden lays out plan to address racial injustice, slams Trump', *Los Angeles Times*.

Hox, Joop J. 2010. *Multilevel Analysis. Techniques and Applications* (Routledge: New York).

Huddy, Leonie. 2013. 'From group identity to political cohesion and commitment.' in, *The Oxford handbook of political psychology, 2nd ed.* (Oxford University Press: New York, NY, US).

Huddy, Leonie, Alexa Bankert, and Caitlin Davies. 2018. 'Expressive Versus Instrumental Partisanship in Multiparty European Systems', *Advances in Political Psychology*, 39: 173-99.

Huddy, Leonie, and Nadia Khatib. 2007. 'American Patriotism, National Identity, and Political Involvement', *American Journal of Political Science*, 51: 63-77.

Jackson, Melinda S. 2011. 'Priming the Sleeping Giant: The Dynamics of Latino Political Identity and Vote Choice', *Political Psychology*, 32: 691-716.

Jacobson, Gary C. 2006. 'The Polls: Polarized Opinion in the States: Partisan Differences in Approval Ratings of Governors, Senators, and George W. Bush', *Presidential Studies Quarterly*, 36: 732-57.

Kam, Cindy D. 2007. 'Implicit Attitudes, Explicit Choices: When Priming Predicts Candidate', *Political Behavior*, 29: 343-67.

Kaplan, Thomas, and Katie Glueck. 2019. 'Joe Biden Calls for Immigration Overhaul, Acknowledging 'Pain' From Deportations', *The New York Times*.

Kellstedt, Paul M., and Guy D. Whitten. 2013. *The Fundamentals of Political Science Research. Second Edition* (Cambridge University Press).

Kelly, Casey Ryan. 2020. 'Donald J. Trump and the Rhetoric of White Ambivalence', *Rhetoric & Public Affairs*, 23: 195-224.

Kelly, Mary Louise, and Casey Morell. 2020. ''The Sheer Volume' Is Hard To Capture: Unemployment In Nevada Soars To Historic High', *NPR*.

Kinder, Donald R., and D. Roderick Kiewiet. 1981. 'Sociotropic Politics: The American Case', *British Journal of Political Science*, 11: 129-61.

King, Jonah B. 2022. 'Divisions in the Big Tent: Group Sentiments and Candidate Preferences within the Democratic Party', *American Politics Research*, 0: 1-15.

Kiousis, Spiro, Jesper Strömback, and Michael McDevitt. 2015. 'Influence of Issue Decision Salience on Vote Choice: Linking Agenda Setting, Priming, and Issue Ownership', *Internation Journal of Communication*, 9: 3347-68.

König, Pascal D., and Thomas Waldvogel. 2022. 'What matters for keeping or losing support in televised debates', *European Journal of Communication*, 37: 312-29.

Layman, Geoffrey C. 1997. 'Religion and Political Behavior in the United States: The Impact of Beliefs, Affiliations, and Commitment From 1980 to 1994', *The Public Opinion Quarterly*, 61: 288-316.

LeBreton, James M., and Jenell L. Senter. 2008. 'Answers to 20 Questions About Interrater Reliability and Interrater Agreement', *Organizational Research Methods 2008*, 11: 815-52.

Lee, Matthew, and Will Weissert. 2020. 'Biden eyes major foreign policy shifts if he wins', *AP News*.

Levendusky, Matthew S. 2017. 'Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization?', *The Journal of Politics*, 80: 59-70.

Lewin, Leif. 1991. *Self-Interest and Public Interest in Western Politics* (Oxford University Press: New York).

Lewis-Beck, Michael S, and Martin Paldam. 2000. 'Economic voting: an introduction', *Electoral studies*, 19: 113-21.

Lewis-Beck, Michael S, and Mary Stegmaier. 2019. 'Economic Voting.' in Roger D. Congleton, Bernard N. Grofman and Stefan Voigt (eds.), *The Oxford Handbook of Public Choice* (OUP USA).

Liebkind, Karmela. 2004. 'Intergroup Relations and Culture.' in Charles D Spielberger (ed.), *Encyclopedia of Applied Psychology* (Elsevier).

Lindsay, James M. 2020. 'The 2020 Election by the Numbers', *Council on Foreign Relations*.

Löfflmann, Georg. 2022. 'Enemies of the People: Donald Trump and the Security Imaginary of America First', *The British Journal of Politics and International Relations*, 24: 543-60.

Lubbers, Marcel, and Marcel Coenders. 2017. 'Nationalistic attitudes and voting for the radical right in Europe', *European Union Politics*, 18: 98-118.

Luttig, Matthew D. 2021. 'Reconsidering the Relationship between Authoritarianism and Republican Support in 2016 and Beyond', *The Journal of Politics*, 83: 783-87.

MacKuen, Michael, and Courtney Brown. 1987. 'Political Context and Attitude Change', *The American Political Science Review*, 81: 471-90.

Malka, Ariel, and Yphtach Lelkes. 2010. 'More than Ideology: Conservative–Liberal Identity and Receptivity to Political Cues', *Social Justice Research*, 23: 156-88.

Mason, Lilliana. 2012. 'The Rise of Uncivil Agreement: Issue Versus Behavioral Polarization in the American Electorate', *American Behavioral Scientist*, 57: 140-59.

———. 2015. '"I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization', *American Journal of Political Science*, 59: 128-45.

———. 2016. 'A Cross-Cutting Calm: How Social Sorting Drives Affective Polarization', *Public Opinion Quarterly*, 80: 351-77.

McCann, Stewart J.H. 2009. 'Political Conservatism, Authoritarianism, and Societal Threat: Voting for Republican Representatives in U.S. Congressional Elections From 1946 to 1992', *The Journal of Psychology*, 143: 341-58.

Miller, Patrick R., and Pamela Johnston Conover. 2015. 'Red and Blue States of Mind: Partisan Hostility and Voting in the United', *Political Research Quarterly*, 68: 225-39.

Moss-Racusin, Corinne, Julie Phelan, and Laurie Rudman. 2010. '"I'm Not Prejudiced, but . . .": Compensatory Egalitarianism in the 2008 Democratic Presidential Primary', *Political Psychology*, 31: 543-61.

Mutz, Diana C. 2002. 'Cross-Cutting Social Networks: Testing Democratic Theory in Practice', *The American Political Science Review*, 96: 111-26.

Nagle, Molly. 2019. 'Biden seeks to undo several Trump immigration policies in first 100 days in new immigration proposal', *abc News*.

Nakagawa, Shinichi, and Holger Schielzeth. 2013. 'A general and simplemethod for obtaining R2 from generalized linear mixed-effects models', *Methods in Ecology and Evolution*, 4: 133-42.

New York Times. 2017. '2016 Presidential Election Results'. https://www.nytimes.com/elections/2016/results/president.

Norrander, Barbara, and Clyde Wilcox. 2008. 'The Gender Gap in Ideology', *Political Behavior*, 30: 503-23.

O'Loughlin, Michael J. 2020. 'Joe Biden's Catholic faith on display at Democratic convention's final night', *America Media*.

Ocampo, Angela X., Sergio I. Garcia-Rios, and Anegla E. Gutierrez. 2021. 'Háblame de tí: Latino mobilization, group dynamics and issue prioritization in the 2020 Election', *The Forum*: 1-28.

Oliver, J. Eric, and Janelle Wong. 2003. 'Intergroup Prejudice in Multiethnic Settings', *American Journal of Political Science*, 47: 567-82.

Pauls, Scott D, Greg Leibon, and Daniel Rockmore. 2015. 'The social identity voting model: Ideology and community structures', *Research & Politics*, 2: 2053168015570415.

Perez, Efren O. 2015. 'Ricochet: How Elite Discourse Politicizes Racial and Ethnie Identities', *Political Behavior*, 37: 155-80.

Powell, G. Bingham. 1976. 'Political Cleavage Structure, Cross-Pressure Process, and Partisanship: An Empirical Test of the Theory', *American Journal of Political Science*, 20: 1-23.

Powell, Richard J. 2004. 'The Strategic Importance of State-Level Factors in Presidential Elections', *Publius*, 34: 115-30.

Putnam, Robert D. 1966. 'Political Attitudes and the Local Community', *The American Political Science Review*, 60: 640-54.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2012a. *Multilevel And Longitudinal Modeling Using Stata. Volume I: Continuous Responses* (Stata Press: Texas, United States).

———. 2012b. *Multilevel And Longitudinal Modeling Using Stata. Volume II: Categorical Responses, Counts, and Survival* (Stata Press: Texas, United States).

Rakich, Nathaniel. 2021. 'How Red Or Blue Is Your State?', *FiveThirtyEight*.

Reher, Stefanie. 2014. "The Effect of Congruence in Policy Priorities on Electoral Participation." In *Elections, Public Opinion and Parties (EPOP)*. Lancaster.

Rhodebeck, Laurie A. 1993. 'The Politics of Greed? Political Preferences among the Elderly', *The Journal of Politics*, 55: 342-64.

Schulman, Bruce J. 2020. 'BU Historian: 2020 Presidential Election a "Momentous Episode in the Story of American Democracy"', *Boston University: Bostonia*.

Sides, John, Michael Tesler, and Lynn Vavreck. 2017. "Donald Trump and the Rise of White Identity Politics." In *The 2016 U.S. Presidential Election: Tumult at Home, Retreat Abroad?* Mershon Center, Ohio State University.

Skelley, Geoffrey. 2019. 'Just How Many Swing Voters Are There?', *FiveThirtyEight*.

Spector, Paul E., and Michael T. Brannick. 2011. 'Methodological Urban Legends: The Misuse of Statistical Control Variables', *Organizational Research Methods*, 14: 287-305.

Stokes-Brown, Atiya Kai. 2006. 'Racial Identity and Latino Vote Choice', *American Politics Research*, 34: 627-52.

Stubager, Rune. 2013. 'The Changing Basis of Party Competition: Education, Authoritarian–Libertarian Values and Voting', *Government and Opposition*, 48: 372-97.

Subramanian, Courtney. 2020. 'Biden's faith comes through in speech, quoting Catholic hymn and Bible verse', *USA Today*.

Swami, Viren, David Barron, Laura Weis, and Adrian Furnham. 2018. 'To Brexit or not to Brexit: The roles of Islamophobia, conspiracist beliefs, and integrated threat in voting intentions for the United Kingdom European Union membership referendum', *British Journal of Psychology*, 109: 156-79.

Tajfel, Henri. 1981. 'Social Categorization, social identity and social comparison.' in, *Human groups and social categories* (Cambridge University Press: Cambridge).

Tajfel, Henri, and J.C. Turner. 1979. 'An integrative theory of intergroup conflict.' in W. G. Austin and S. Worchel (eds.), *The social psychology of intergroup relations* (Brooks/Cole: Monterey, CA).

Transue, John E. 2007. 'Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force', *American Journal of Political Science*, 51: 78-91.

Turner, John C. 1987. 'Rediscovering the social group: A self-categorization theory.' in, *Rediscovering the social group: A self-categorization theory.* (Basil Blackwell: Cambridge, MA, US).

U.S. Bureau of Labor Statistics. 2022. 'Local Area Unemployment Statistics', Accessed 19.09.2022. https://www.bls.gov/lau/lastrk20.htm.

United States Census Bureau. 2021. 'Racial and Ethnic Diversity in the United States: 2010 Census and 2020 Census', Accessed 19.09.2022. https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html.

———. 2022. 'Quick Facts', Accessed 19.09.2022. https://www.census.gov/quickfacts/fact/table/US/PST045221.

van der Eijk, Cees. 2002. 'Design issues in electoral research: taking care of (core) business', *Electoral studies*, 21: 189-206.

Westreich, Daniel, and Sander Greenland. 2012. 'The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients', *American Journal of Epidemiology*, 177: 292-98.

Whitehead, Andrew L., Samuel L. Perry, and Joseph O. Baker. 2018. 'Make America Christian Again: Christian Nationalism and Voting for Donald Trump in the 2016 Presidential Election', *Sociology of Religion: A Quarterly Review*, 79: 147-71.

Wilms, R., E. Mäthner, L. Winnen, and R. Lanwehr. 2021. 'Omitted variable bias: A threat to estimating causal relationships', *Methods in Psychology*, 5: 1-10.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion* (Cambridge Unviersity Press).

Ziller, Conrad. 2018. 'Multiple Regression mit voneinander abhängigen Beobachtungen.' in C. Wagemann, A. Goerres and M. Siewert (eds.), *Handbuch Methoden der Politikwissenschaft* (Springer VS: Wiesbaden).

Zingher, Joshua, and Michael E. Flynn. 2019. 'Does Polarization Affect Even the Inattentive? Assessing the Relationship Between Political Sophistication, Policy Orientations, and Elite Cues', *Electoral studies*, 57: 131-42.

Appendix

APPENDIX 1: Table of Number of Observations by State

Table 1: Number of Observations by State

State	Number of Observations	State	Number of Observations
Alabama	63	Montana	17
Alaska	6	Nebraska	38
Arizona	102	Nevada	24
Arkansas	31	New Hampshire	35
California	338	New Jersey	95
Colorado	108	New Mexico	27
Connecticut	43	New York	209
Delaware	9	North Carolina	170
Florida	252	North Dakota	14
Georgia	111	Ohio	199
Hawaii	9	Oklahoma	59
Idaho	45	Oregon	72
Illinois	176	Pennsylvania	229
Indiana	128	Rhode Island	12
Iowa	48	South Carolina	68
Kansas	71	South Dakota	9
Kentucky	81	Tennessee	109
Louisiana	49	Texas	250
Maine	27	Utah	56
Maryland	95	Vermont	13
Massachusetts	128	Virginia	103
Michigan	182	Washington	133
Minnesota	127	West Virginia	33
Mississippi	36	Wisconsin	118
Missouri	110	Wyoming	7

APPENDIX 2: Original Wording of Questions in ANES 2020 Survey

Table 2: Question-Wording of all Survey Variables Used from the ANES 2020 Time Series Study as Provided in the Codebook

Variable	Question Wording
Feeling Thermometer: Joe Biden	*I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.* *How would you rate:* *Joe Biden*
Feeling Thermometer: Donald Trump	*I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.* *How would you rate:* *Donald Trump*
White Identity	*How important is being white to your identity?*
National Identity	*How important is being American to your identity?*
Religious Identity	*What is your present religion, if any?*
National Economy	*Now thinking about the economy in the country as a whole, would you say that over the past year the nation's economy has gotten better, stayed about the same, or gotten worse?*
Education	*What is the highest level of school you have completed or the highest degree you have received?*
Gender	*What is your sex?*

APPENDIX 3: Ethnic Diversity Index

Diversity Index Equation as calculated by the United States Census Bureau (Source: https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html)

Based on:

Peter Michael Blau, *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, New York: Free Press, Vol. 7, 1977.

Phillip Meyer and Shawn McIntosh, "The USA Today Index of Ethnic Diversity," *International Journal of Public Opinion Research*, Volume 4, Issue 1, 1992, pp. 51–58.

(Equation 1)

$$DI = 1 - (H^2 + W^2 + B^2 + AIAN^2 + Asian^2 + NHPI^2 + SOR^2 + Multi^2)$$

- ***H*** is the proportion of the population who are Hispanic or Latino.
- ***W*** is the proportion of the population who are White alone, not Hispanic or Latino.
- ***B*** is the proportion of the population who are Black or African American alone, not Hispanic or Latino.
- ***AIAN*** is the proportion of the population who are American Indian and Alaska Native alone, not Hispanic or Latino.
- ***Asian*** is the proportion of the population who are Asian alone, not Hispanic or Latino.
- ***NHPI*** is the proportion of the population who are Native Hawaiian and Other Pacific Islander alone, not Hispanic or Latino.
- ***SOR*** is the proportion of the population who are Some Other Race alone, not Hispanic or Latino.
- ***MULTI*** is the proportion of the population who are Two or More Races, not Hispanic or Latino.

APPENDIX 4: Residual Diagnostics

Figure 1: Histogram of Level-1 Residuals with Overlayed Normal Distribution Overlay

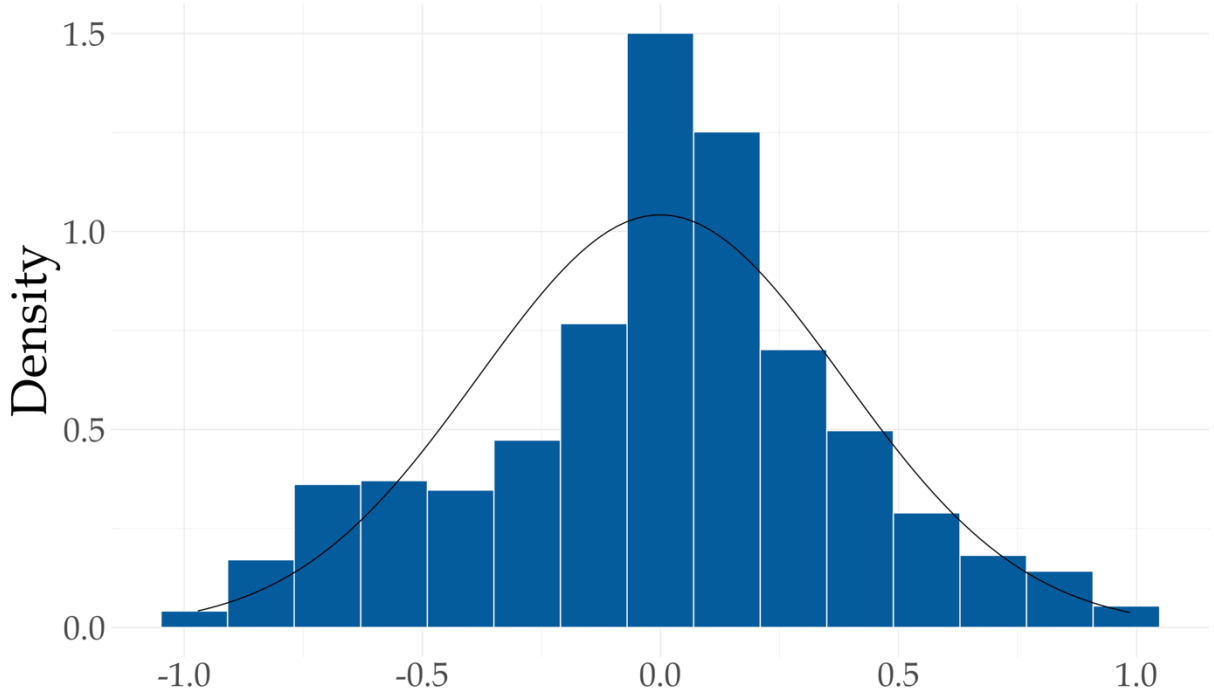
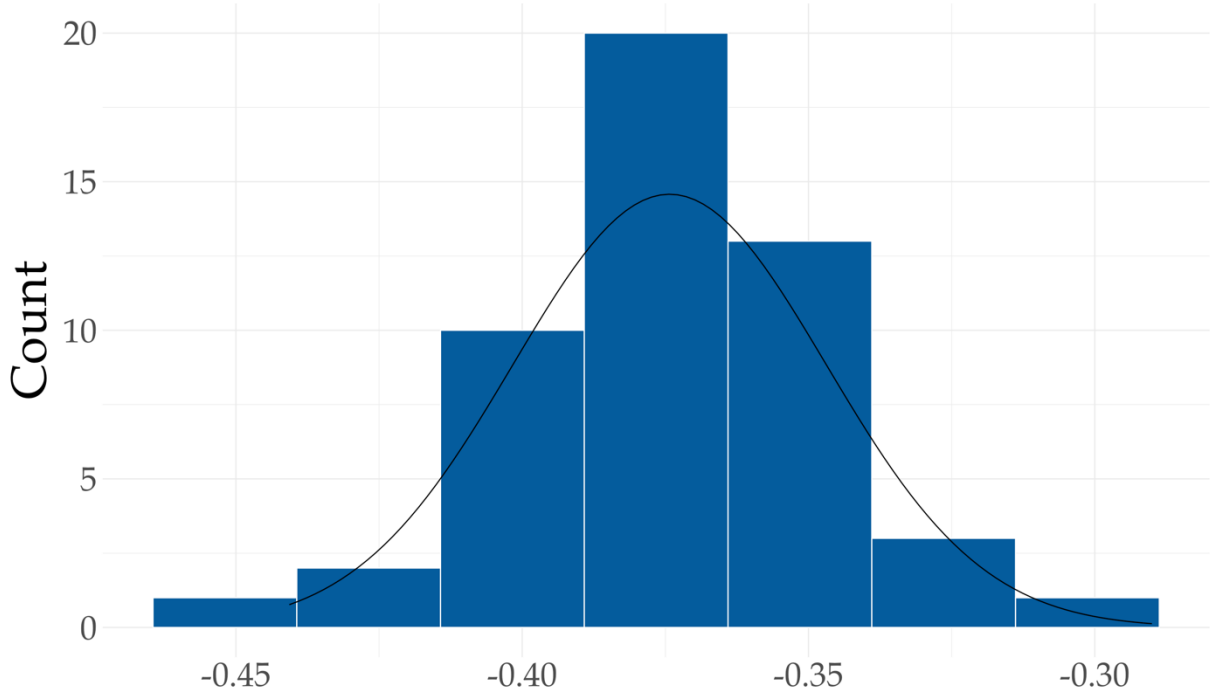


Figure 2: Histogram of Random Intercepts with Normal Distribution Overlay



APPENDIX 5: Additional Visualizations for Bivariate Analysis

Figure 3: Bivariate Bar Plot of White Identity Grouped by Candidate Preference

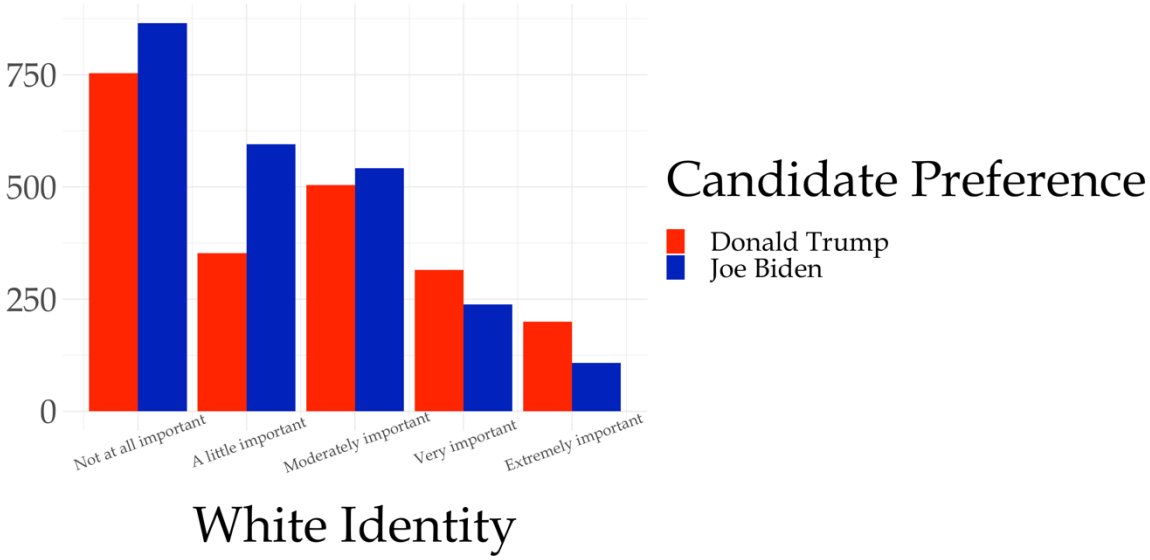


Figure 4: Bivariate Bar Plot of White Identity Grouped by Candidate Preference

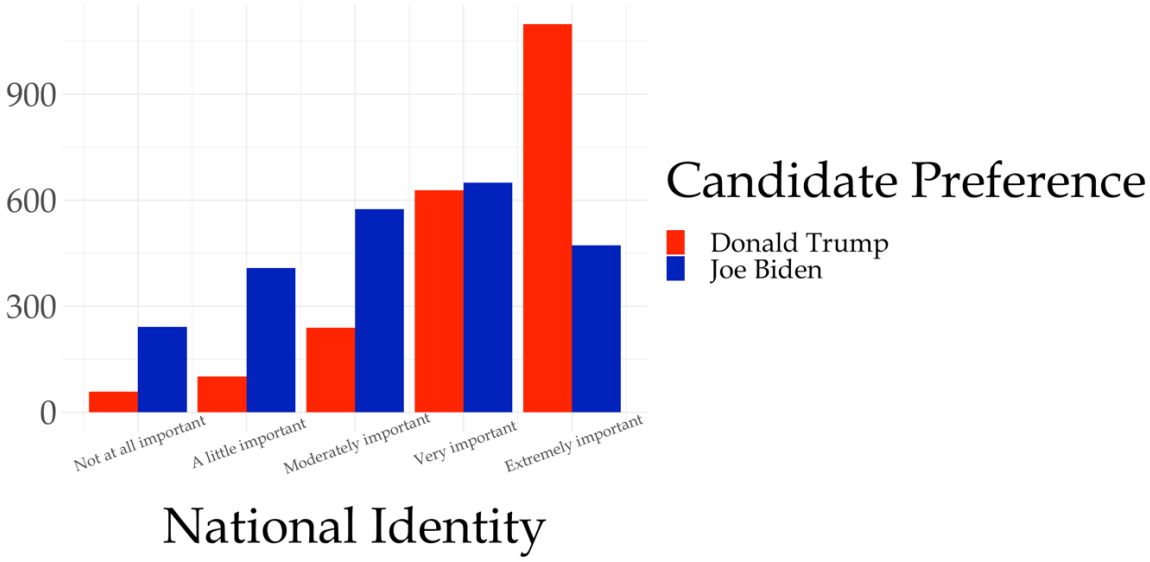


Figure 5: Bivariate Scatterplot of Ethnic Diversity in Each State and Share of People Preferring Joe Biden

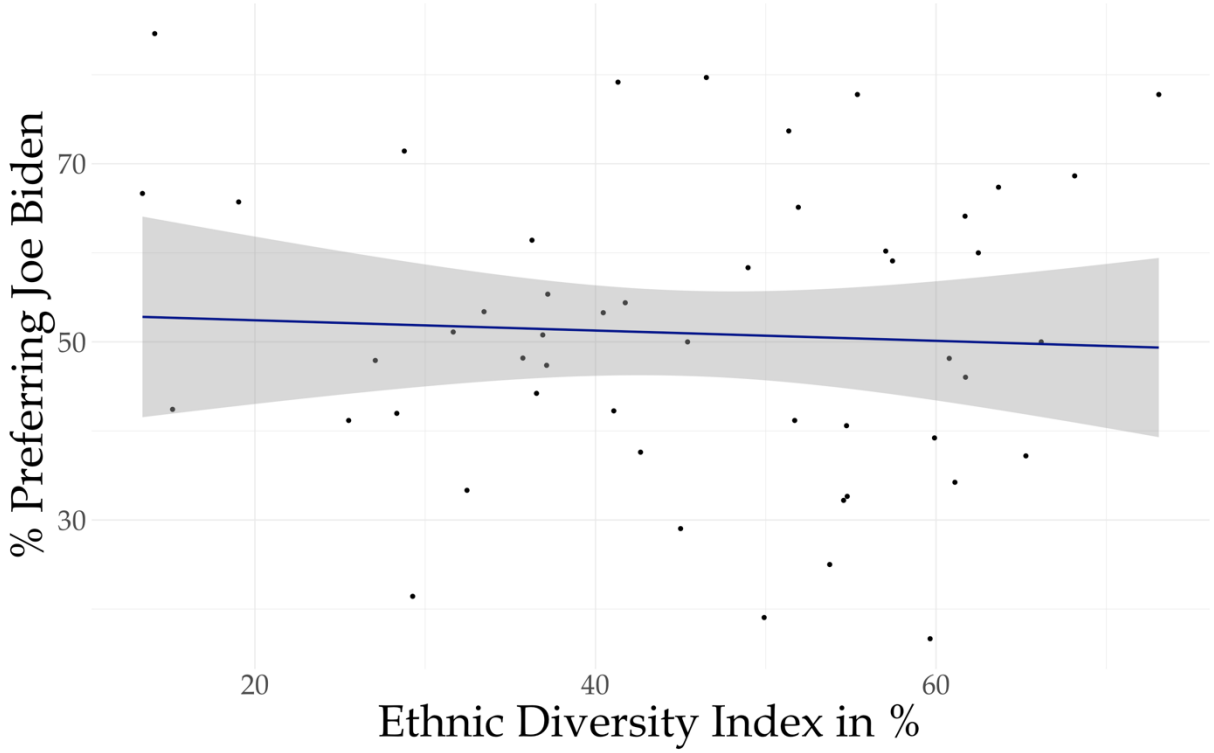


Figure 6: Bivariate Scatterplot of Unemployment Rate in Each State and Share of People Preferring Joe Biden

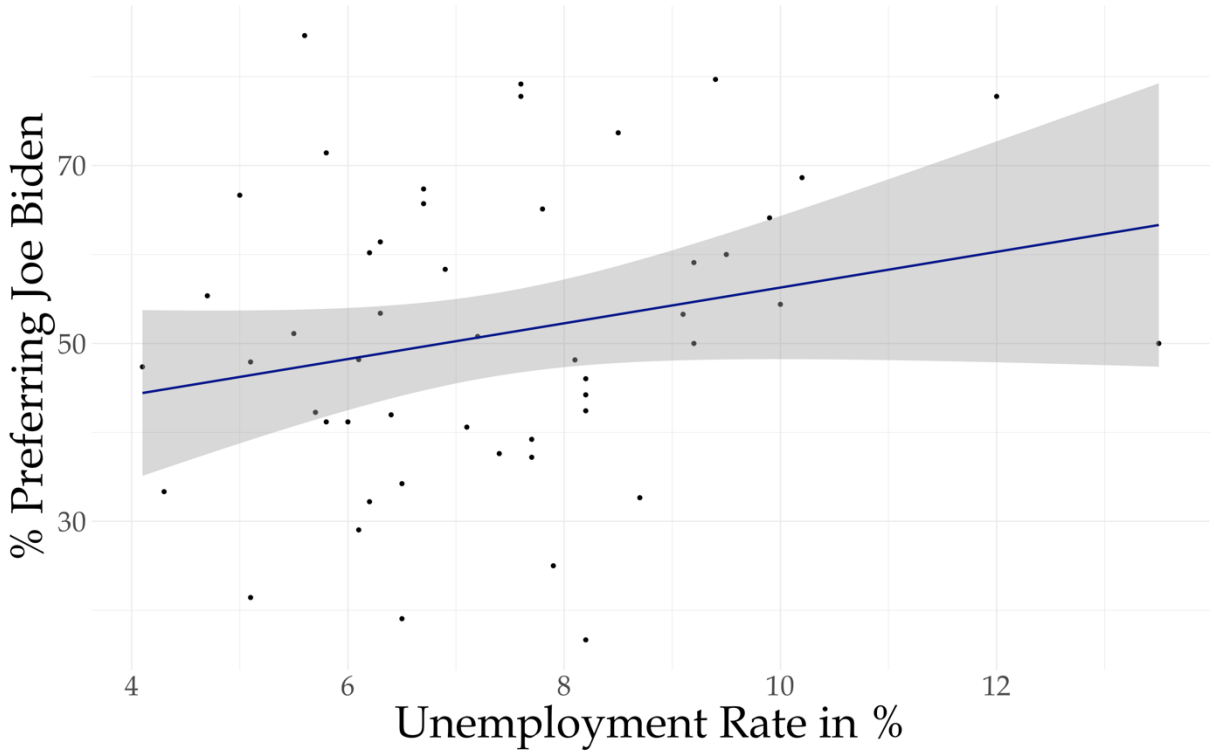


Figure 7: Jittered Boxplot for Share of People Preferring Joe Biden by Partisan Lean in Each State

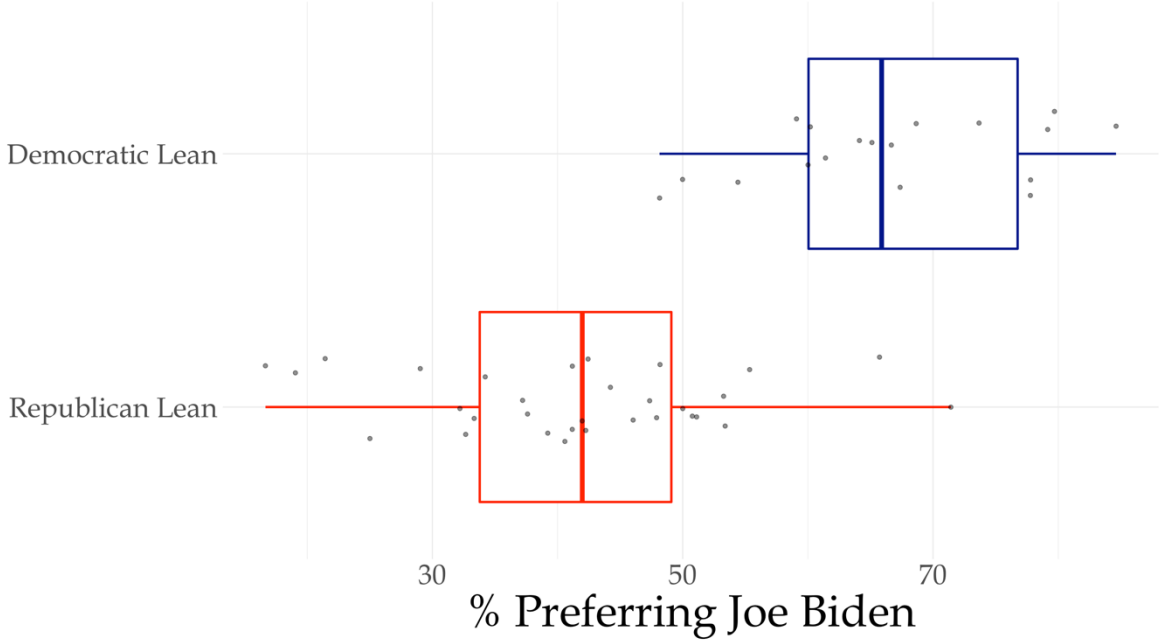


Figure 8: Bivariate Scatterplot of Ethnic Diversity in Each State and Share of People Preferring Donald Trump

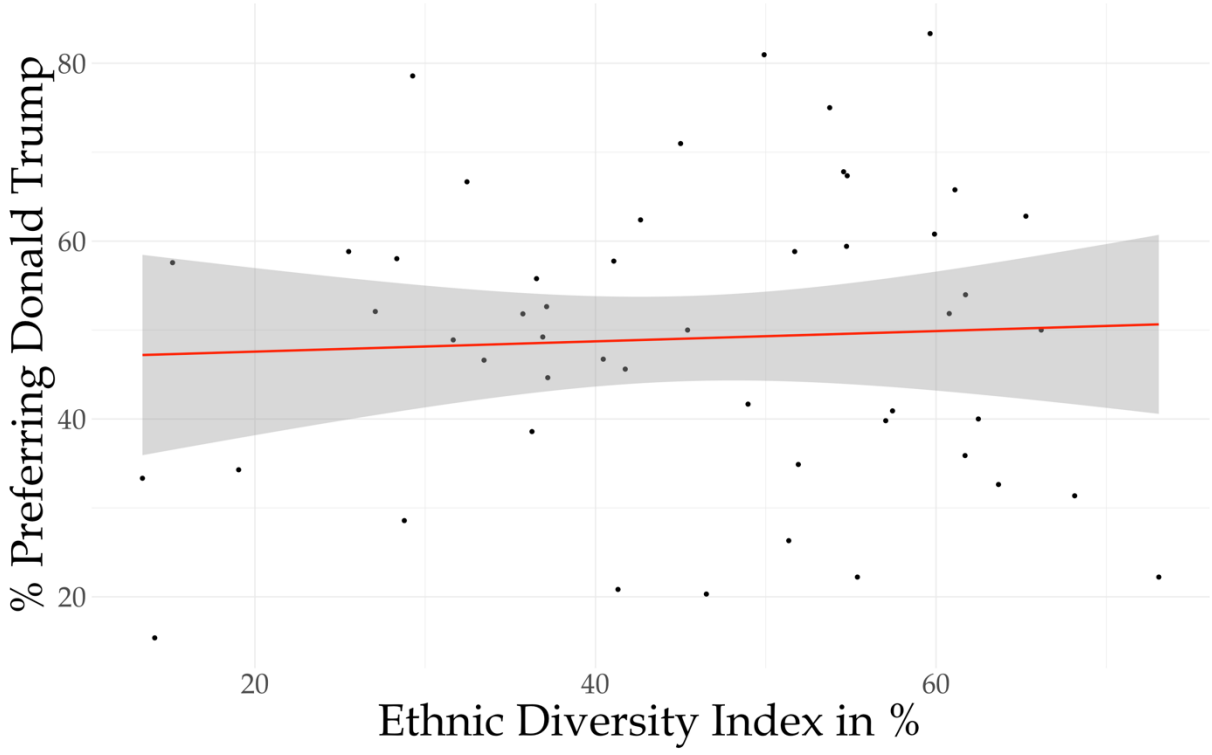


Figure 9: Bivariate Scatterplot of Unemployment Rate in Each State and Share of People Preferring Donald Trump

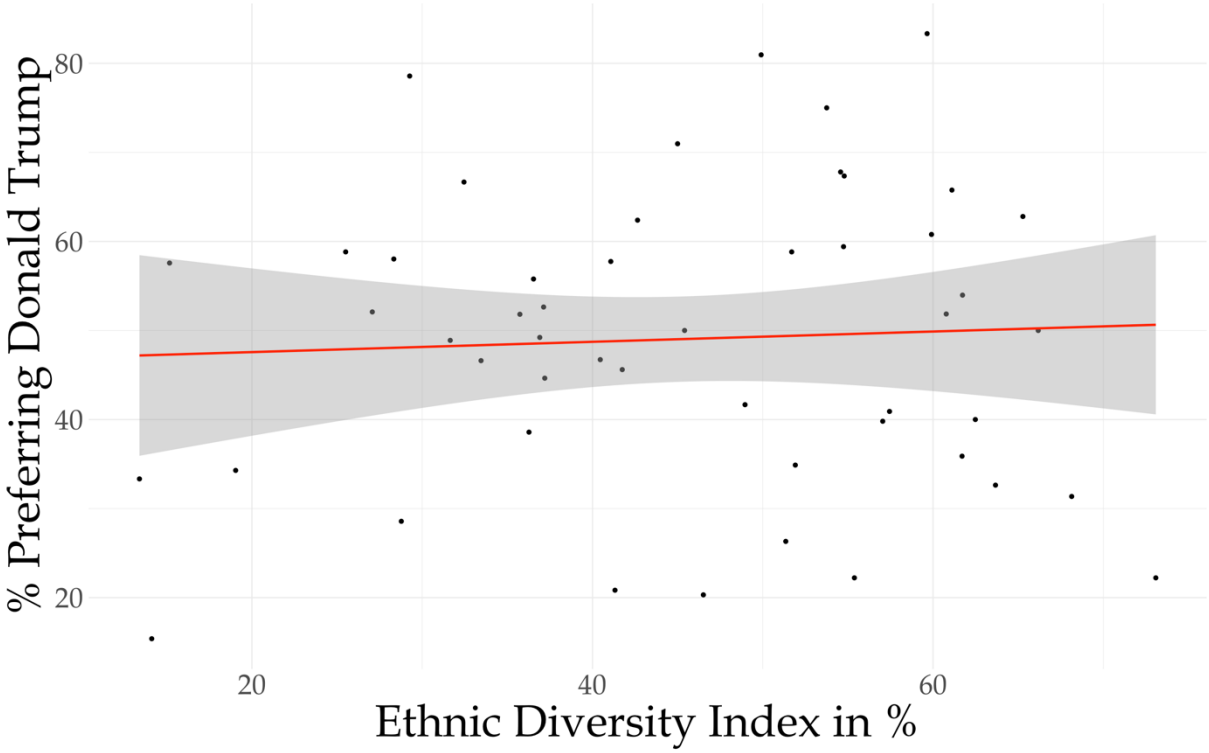
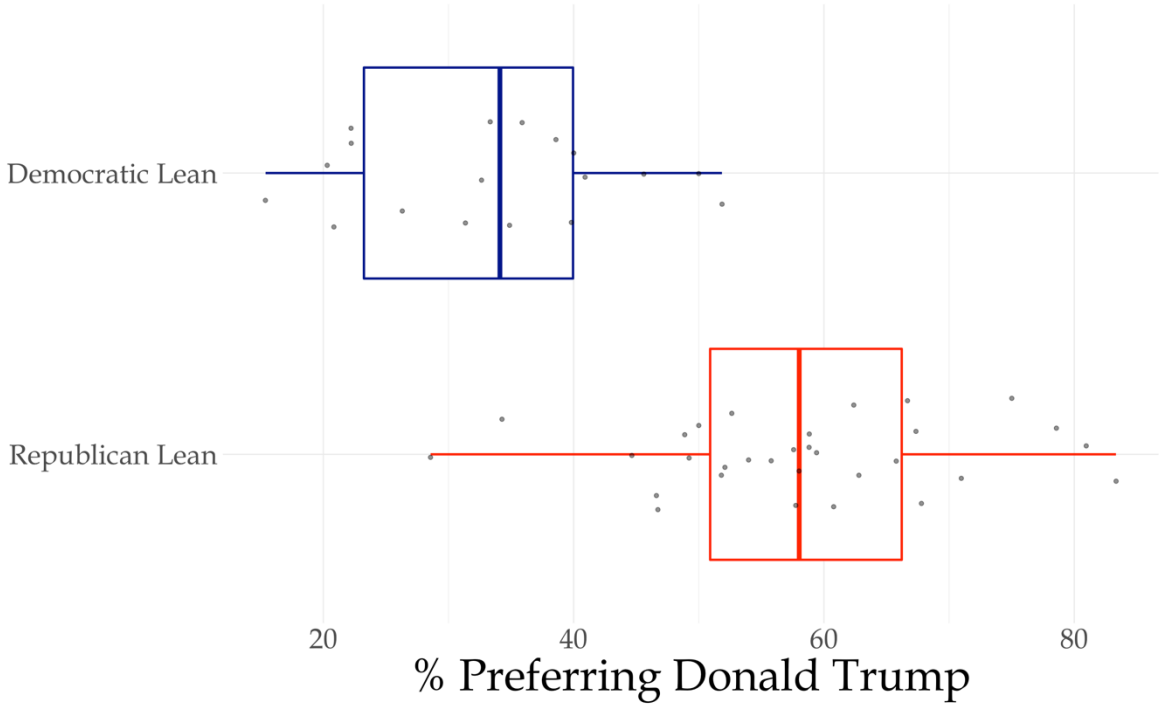


Figure 10: Jittered Boxplot for Share of People Preferring Donald Trump by Partisan Lean in Each State



APPENDIX 6: Likelihood Ratio Test

Table 3: Results of Likelihood Ratio Test Between the Individual Model and Full Model for Model Comparison

Model	Log Likelihood	Degrees of Freedom	Chi²-value	p-value
Individual Model	*-2030.37*			
Full Model	*-2013.23*	*4*	*34.27*	*0.00*

APPENDIX 7: Robustness Checks

Table 4: Multilevel Linear Regression Models for the Continuous Dependent Variables of Feeling Thermometer Toward the Two Candidates

	DV: Thermometer Joe Biden		**DV: Thermometer Donald Trump**	
	Coefficient	*Std. Error*	*Coefficient*	*Std. Error*
Intercept	34.96 ***	1.66	53.71 ***	1.78
Individual Level				
White Identity	1.75 ***	0.47	-1.37 **	0.51
National Identity	-5.77 ***	0.50	9.61 ***	0.55
Religious Identity: Ref. Not Religious				
Evangelical	-18.00 ***	1.55	27.37 ***	1.70
Other Christian	-6.95 ***	1.14	15.19 ***	1.25
Other Religion	4.40 *	2.00	-0.34	2.19
National Economy: Ref. Stayed the same				
Gotten Better	-14.54 ***	1.40	17.29 ***	1.53
Gotten Worse	20.59 ***	1.16	-26.36 ***	1.27
State Level				
Ethnic Diversity	-0.83	0.66	1.52 *	0.66
Unemployment Rate	0.42	0.70	-0.47	0.70
Partisan Lean				
Democratic	7.35 ***	1.37	-8.19 ***	1.36
Neutral	1.00	3.69	-3.48	3.66
Control Variables	Yes		Yes	
Random Effects				
σ^2	828.72		991.35	
τ_{00}	5.24 $_{state}$		3.53 $_{state}$	
ICC	0.01		0.00	
N	50 $_{state}$		50 $_{state}$	
Observations	4474		4474	
Marginal R^2 / Conditional R^2	0.342 / 0.346		0.450 / 0.452	
BIC	42877.59		43670.28	

** p<0.05 ** p<0.01 *** p<0.001*

APPENDIX 8: Interaction Between White Identity and National Identity

Table 5: Regression Table of Full Model Including Interaction Effect Between White Identity and National Identity

Dependent Variable: Favoring Donald Trump (0) or Joe Biden (1)

	Interaction Model	
	Logit	*Std. Error*
Intercept	-0.36 *	0.14
Individual Level		
White Identity	0.15 **	0.05
National Identity	-0.72 ***	0.05
Religious Identity: Ref. Not Religious		
Evangelical	-2.04 ***	0.15
Other Christian	-1.07 ***	0.11
Other Religion	0.14	0.22
National Economy: Ref. Stayed the same		
Gotten Better	-1.43 ***	0.15
Gotten Worse	1.53 ***	0.10
State Level		
Ethnic Diversity	-0.13 **	0.05
Unemployment Rate	0.02	0.05
Partisan Lean: Ref. Republican		
Democratic	0.64 ***	0.11
Neutral	0.17	0.27
Interaction		
White Identity x National Identity	-0.04	0.05

Control Variables	Yes
Random Effects	
σ^2	3.29
$\tau_{00\ state}$	0.01
ICC	0.00
$N_{\ state}$	50
Observations	4474
Marginal R^2 / Conditional R^2	0.540 / 0.541
AIC	4057.994

** p<0.05 ** p<0.01 *** p<0.001*

Table 6: Pearson's Correlation Coefficient for White Identity and National Identity

	Pearson's R	**P-value**	**95% Upper Bound**	**95% Lower Bound**
White Identity x National Identity	*0.39*	*0.00*	*0.41*	*0.36*

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

FAKULTÄT FÜR GESELLSCHAFTSWISSENSCHAFTEN

INSTITUT FÜR POLITIKWISSENSCHAFT

Versicherung an Eides Statt

Name: Diaz Garcia, Manuel

Ich versichere durch meine eigenhändige Unterschrift an Eides Statt, dass ich

die vorliegende Hausarbeit mit dem Titel

The Social Identities of Biden and Trump Sympathizers – A Multilevel-Analysis of the Candidate Preference in the U.S. Presidential Election 2020

selbstständig und ohne fremde Hilfe angefertigt habe und alle Stellen, die ich wörtlich oder dem Sinne nach aus veröffentlichten und nicht veröffentlichten Quellen entnommen habe, als solche kenntlich gemacht habe und mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht zur Bewertung vorgelegt worden ist.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs.1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Duisburg, 10.10.2022

Ort, Datum

Unterschrift