American Psychological Association
American Psychology-Law Society

Law and Human Behavior

ISSN: 0147-7307

2025, Vol. 49, No. 3, 250–262
https://doi.org/10.1037/lhb0000600

Experts, Commercial Software, and the Internal Revenue Service: American Taxpayer Perceptions of Trust and Procedural Justice

Krystia Reed[1], Morgan Wagner[1], Saeid Tizpaz-Niari[2], and Ashutosh Trivedi[3]
[1] Department of Psychology, University of Texas, El Paso
[2] Department of Computer Science, University of Texas, El Paso
[3] Department of Computer Science, University of Colorado Boulder

Objective: The complexity of tax laws makes manual preparation difficult, leading more taxpayers to use software or accountants. This study presents an experimental analysis comparing taxpayer perceptions of trust and procedural justice when filing with tax experts versus using tax software. The study addressed four questions: (1) How do perceptions of human tax experts compare to tax software? (2) Do perceptions vary among different types of tax software? (3) Do trust and procedural justice predict filing decisions? (4) Can taxpayers effectively oversee tax preparation software? ***Hypotheses:*** We hypothesized that participants would favor professional tax experts over commercial and Internal Revenue Service (IRS) software (Hypothesis 1). We also expected higher procedural justice to correlate with greater satisfaction (Hypothesis 2), self-identified knowledge to correlate with accurate expectations (Hypothesis 3), and filing decisions to be predicted by trust in the method (Hypothesis 4a) or outcome (Hypothesis 4b). We anticipated that higher trust in software would increase the likelihood of filing with software (Hypothesis 5) and that inconsistency across methods would decrease filing likelihood (Hypothesis 6). ***Method:*** In the experiment, 146 taxpayers (64% women; 88% Hispanic; 75% White; M_{age} = 27.55 years) prepared their taxes using three methods: a tax professional, commercial software, and IRS software. Participants rated and ranked the trustworthiness of each method and indicated their preference. ***Results:*** As predicted, participants had the most favorable perceptions of the tax expert, followed by commercial software and IRS software (Hypothesis 1). Trust in the method, not the outcome, predicted filing decisions (Hypothesis 4a). Participants with higher trust in software were more likely to file with software (Hypothesis 5). Contrary to expectations, procedural justice did not correlate with satisfaction (Hypothesis 2), and knowledge did not correlate with accurate expectations (Hypothesis 3). Consistency across methods did not predict filing (Hypothesis 6). ***Conclusions:*** Participants generally preferred human experts, but trust in software could override this preference. Future research directions and implications are discussed.

Public Significance Statement
Lower income taxpayers file using methods they find more trustworthy when costs are equal but might default to a cheaper option when costs of the trustworthy option increase. Research should be done on how free question-answer sessions help low-income taxpayers in successfully using free tax filing options. In addition, research should expand to examine filing behavior of higher income taxpayers.

Keywords: tax law, tax experts, software, procedural justice

This article was published Online First May 15, 2025.
David DeMatteo served as Action Editor.
Krystia Reed https://orcid.org/0000-0003-4373-3450
Morgan Wagner https://orcid.org/0000-0001-8422-6979
Saeid Tizpaz-Niari https://orcid.org/0000-0002-1375-3154
Ashutosh Trivedi https://orcid.org/0000-0001-9346-0126
Saeid Tizpaz-Niari is now at the Department of Computer Science, University of Illinois Chicago.

Hypotheses were pre-registered and materials are available on the Open Science Framework (https://osf.io/emrxs/). The authors are committed to transparency and open science. The hypotheses have been preregistered and all materials are stored on the Center for Open Science website. However, due to the confidential nature of tax information, even with de-identification there are significant concerns about making these data publicly available. Therefore, for the privacy of our participants, the data are not publicly available. This data has not been presented previously.

This project was funded by a grant from the Designing Accountable Software Systems program at National Science Foundation (Awards CCF-2317206 and CCF-237207) to Saeid Tizpaz-Niari, Krystia Reed, and Ashutosh Trivedi. The authors would like to thank the research assistants necessary to conduct this deliberation study (in alphabetical order): Noe Aguilar, Brianna Bustillos, Brissa Cervantes, William Fenstermacher, Chelsy Garcia, Karyme Garcia, Katia Gutierrez, Grace Hanzelin, Judith Hinojosa, Sarah Maul, Arisa Morales, Mayew Quijas, Kimberley Rodriguez Ramirez, Benjamin Sanchez Jimenez, Jennifer Santaella, Emiliano Silva, Lilia Tovar, Ashley Vaughan, Arianna Villalobos, and Crystal Walls.

Krystia Reed played a lead role in conceptualization, data curation, investigation, methodology, project administration, resources, software, supervision, writing–original draft, and writing–review and editing, a supporting role in formal analysis, and an equal role in funding acquisition and

continued

As Hamilton (1787) wrote in the Federalist Papers,

> there must be interwoven, in the frame of the government, a general power of taxation. … A complete power … to procure a regular and adequate supply of [money] … may be regarded as an indispensable ingredient in every constitution. (para. 1–2)

Paying taxes is not only a civic duty but also a crucial means of ensuring national security and preserving the freedoms we enjoy. As shown in Figure 1, an average of 135.6 million Americans filed tax returns annually between 2008 and 2023 (Internal Revenue Service [IRS], 2024a). However, 5% of American taxpayers fail to file their taxes, which can result in significant penalties for the taxpayer and impact governmental operations (Picchi, 2023). One predictor of tax compliance worldwide is trust in the tax authority (Nurkholis et al., 2020), but research has yet to examine how trust in tax software specifically influences filing decisions.

As tax filing becomes increasingly intertwined with technology, the landscape of tax compliance is rapidly evolving. During the past 16 years, the rate of electronic filing (hereafter, e-filing) has surged by 46%, with 95% of returns being e-filed in 2023 (see Figure 1). This growth includes a 20% increase in e-filing by tax professionals and an impressive 94% rise among individual taxpayers, many of whom now file without expert assistance. This shift is likely driven by the growing reliance on commercial tax software (Collins, 2024). Considering this increasing intersection between taxes and technology, we explored in the present study how trust in tax software shapes taxpayer perceptions and behaviors.

This article begins by reviewing existing research on trust and procedural justice in two key areas: tax compliance and technology/software. Following this, we present the results of an experiment that compared taxpayer perceptions and behaviors across three different tax preparation methods: using an expert, commercial software, and Internal Revenue Service (IRS) software. We conclude by discussing the implications of these findings for future research and tax policy.

Trust and Procedural Justice With Tax Authorities

Within institutions, the process-based judgments of trust and procedural justice drive perceptions of legitimacy, cooperation and compliance, and acceptance of decisions (Tyler, 2003). When people trust the motives of the system (*trust*) and perceive the process as fair (*procedural justice*), they are more likely to follow the law (Tyler, 1990). Notably, research on procedural justice in the legal system has found that perception of a fair process is a stronger predictor of overall satisfaction than the outcome itself, for both criminal defendants (Landis & Goodstein, 1986; Tyler, 1984) and civil defendants (Shestowsky, 2014). Whereas much of the research on legal trust and procedural justice has concentrated on the justice system (Tyler, 1984) and policing (Tyler, 2011), findings also indicate that procedural justice is relevant to the perceptions of legitimacy of government more broadly (Tyler, 1994).

One governmental institution in which legitimacy might be particularly important is the tax system, which in the United States is the IRS. As Hamilton noted in the Federalist Papers, taxes are essential to a functioning government. By collecting money from its citizens, governments can fund essential services, such as schools, physical infrastructure (e.g., highways), defense, public safety (e.g., police, firefighters), the justice system, and other public services (IRS, n.d.). Consequently, it is extremely important for a functioning government to ensure that citizens pay taxes. In 2023 alone, the IRS spent more than half a million dollars on enforcement of tax collections (IRS, 2024c). If people do not pay taxes, there can be significant penalties:

> Any person who willfully attempts … to evade or defeat any tax imposed … shall, in addition to other penalties provided by law, be guilty of a felony and, upon conviction thereof, shall be fined not more than $100,000 ($500,000 in the case of a corporation), or imprisoned not more than 5 years, or both. (*Attempt to Evade or Defeat Tax, 1954*; https://www.law.cornell.edu/uscode/text/26/7201)

However, despite the wide applicability of tax law, the governmental implications of having a successful tax system, and the major penalties for not complying with tax law, limited research has examined taxpayer behavior within the United States. Research conducted internationally suggests that perceptions of trust and procedural justice can increase tax compliance, but it is unclear how these perceptions differ on the basis of whether the taxes were prepared by a human professional or using tax software.

Tax Compliance

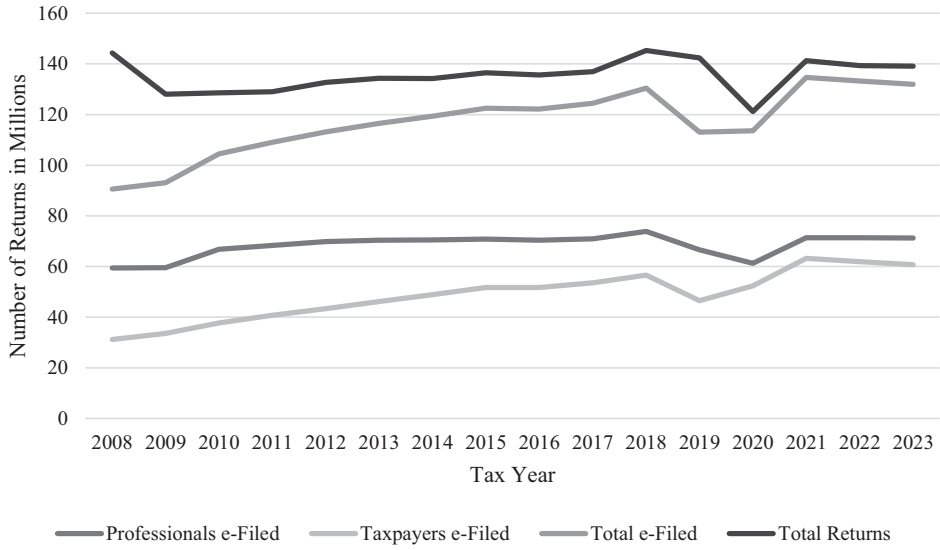
Internationally, higher perceptions of trust and procedural justice have been linked to increased taxpayer compliance. A meta-analysis of 102 studies worldwide found that higher perceptions of procedural justice and trust with tax authorities resulted in increased tax compliance (Nurkholis et al., 2020). Taxpayers do not even need to have direct experience with the tax authority; merely telling people that there is a climate of trust results in greater support—and telling them that there is a lack of trust results in less support—of hypothetical tax policies (Gangl et al., 2020). Moreover, higher perceptions of procedural justice and legitimacy of the tax authority have been found to predict later tax compliance for Australian taxpayers who had previously been sanctioned with tax avoidance (Murphy et al., 2016).

Although research has not directly examined tax behavior in the United States, some research has compared hypothetical perceptions of taxes between the United States and other countries. For example, Americans have been found to have higher willingness to pay taxes compared with 14 European countries, a finding that is based in the United States being more transparent than other countries (Alm & Torgler, 2006). D'Attoma (2020) experimentally examined this

visualization. Morgan Wagner played a lead role in formal analysis, a supporting role in conceptualization, writing–original draft, and writing–review and editing, and an equal role in data curation, investigation, methodology, project administration, supervision, and visualization. Saeid Tizpaz-Niari played a supporting role in investigation, methodology, and writing–review and editing and an equal role in funding acquisition. Ashutosh Trivedi played a supporting role in investigation, methodology, and writing–review and editing and an equal role in funding acquisition.

Correspondence concerning this article should be addressed to Krystia Reed, Department of Psychology, University of Texas, El Paso, 500 West University Avenue, El Paso, TX 79902, United States. Email: kmreed2@utep.edu

Figure 1
Tax Filing Trends 2008–2023



Note. Data for this graph came from the *Free File: About the Free File Alliance*, by Internal Revenue Service, 2024b (https://www.irs.gov/e-file-providers/about-the-free-file-alliance).

perception by asking American and Italian participants to pay money to different institutions. When the institutions were identical, both groups complied at the same rate; however, when participants were asked to pay money to their tax institution, Italians were less compliant than Americans. These differences were attributed to Italians showing less trust in government than did Americans (D'Attoma, 2020). In another study comparing Americans with Ethiopians, researchers found that perceptions of procedural justice were more important in predicting Americans' compliance with tax law than Ethiopians' (van Dijke et al., 2019). Consequently, perceptions of trust and procedural justice appear to be linked to Americans' hypothetical tax behavior, but research has yet to examine how they influence tax preparation.

Tax Software

Given the increased use of tax software, it is important to consider whether perceptions of trust, procedural justice, and legitimacy vary when technology-based preparation methods rather than human tax preparers are used. General research on technology usage suggests that people are more likely to use software if they trust the software and technology (Mcknight et al., 2011; Söllner et al., 2016). These findings apply across a variety of domains (education: Gulati et al., 2019; online purchases: ElSayad & Mamdouh, 2024; COVID-19 tracing apps: Sousa & Kalju, 2022). In fact, higher trust in technology has been linked to more willingness to e-file taxes (Carter et al., 2011; Schaupp & Carter, 2010).

Beyond trust in technology, perceptions of risk have also been found to be an important factor when it comes to tax filing specifically. One study found that trust in e-filing was linked to higher willingness to use e-file, but the effect was moderated by risk tolerance (Akram et al., 2019). Conversely, there is also evidence that people fear using online tax software because of a perceived security risk, but the effect is moderated by trust and procedural justice (Chu et al., 2022; McLeod et al., 2009). Concerns about risk might be valid—tax software is prone to hacking, and errors using e-file might be judged more harshly by the IRS (Starkman, 2024). Beyond security, there might be valid concerns that software is prone to errors (Tizpaz-Niari et al., 2023).

It is also important to note that most research has focused on perceptions of e-filing and not tax software itself, and e-filing and tax software are not synonymous. Many accountants prepare tax returns and e-file for their clients, and taxpayers can prepare their returns using software without e-filing (there is sometimes even an additional fee to e-file). Thus, it is important to consider perceptions of the software beyond perceptions of e-filing.

Human Experts Versus Software

Tax preparation can be a complicated process that involves applying the newest law to the taxpayer's specific situation. Many taxpayers cannot prepare their own taxes and, therefore, rely on assistance from either a human expert or software. There is no research directly comparing how tax experts are perceived relative to tax software; however, general research on perceptions of human and technology decision making can be instructive. One study found that citizens rated decisions as more trustworthy when made by a human than by software (Ingrams et al., 2022), which could be important for taxes. However, most research on this topic has been conducted in the context of employment. Although employment decisions are not directly relevant to tax decisions, they both usually involve a complicated process that results in a higher authority (employer or IRS) imposing a decision on the person (employee or taxpayer) through an agent (manager or tax preparer). Thus, the findings in this domain may offer guidance when hypotheses about perceptions of different types of tax preparers are developed.

Unfortunately, research comparing different types of decision makers in this domain is mixed.

Some research indicates that computer agents are perceived as fairer and more trustworthy than human agents. For example, participants rated computer evaluations of job candidates as more trustworthy, fair, and useful than human evaluations (Choung et al., 2024). This might be because of general perceptions that computers are less emotional and more data driven than humans (Helberger et al., 2020). Other research suggests the opposite (Abric & Kahan, 1972; Acikgoz et al., 2020; Liu, 2021). Participants rated employment interviews as fairer and more procedurally just when conducted by a human than a computer (Acikgoz et al., 2020). And still other research has found no differences: In a series of two studies, Ötting and Maier (2018) looked at German employee perceptions of organizational allocation decisions (new tasks or vocational trainings) and found no significant effects of their manipulated decision agent (i.e., the entity delivering the decision—human, computer, or robot). Thus, the relationship between agent and fairness is complicated.

These differences might be partially explained by the complexity of the decision. Nagtegaal (2021) manipulated the decision maker and complexity of decisions in two employment studies. For low complexity decisions, employees had higher perceptions of procedural justice when the decision was made by a computer algorithm compared with a human; for high complexity decisions, employees had lower perceptions of procedural justice when the decision was made by a computer algorithm compared with a human (Nagtegaal, 2021). Thus, it is possible that humans are perceived more positively for high complexity tasks.

Research on social influence might also be important to consider. Trust has been found to be affected by social learning and influence (Simonsen et al., 2014; Wei et al., 2019). A meta-analysis of 32 studies indicates that, even when interacting virtually, participants demonstrate more social influence when interacting with other humans than computers (Fox et al., 2015). In fact, social influence is indirectly linked to whether people shift from using human bankers to banking software, with people more likely to use the human if they perceive greater social influence (Chaouali et al., 2016). Thus, it is likely that there will be some social influence from a human tax preparer that might result in more trust in the human compared with the software. There is also some evidence that more interactive software is perceived more favorably than less interactive software (Blascovich et al., 2002; Nowak & Fox, 2018), possibly because of social influence.

Perceptions of technology and software use may also be influenced by individual differences. When people are less familiar with technology, it is more likely to be perceived as an "outgroup member" that is less trustworthy; increased familiarity is linked to increases in trustworthiness and decreases in uncertainty (Liu, 2021). People with more access to technology also tend to trust technology more (Araujo et al., 2023).

Taken together, the existing research suggests that human tax preparers will likely be perceived as more trustworthy than tax software. Tax preparation is generally perceived to be a complicated process, which research shows results in better perceptions of humans than software. Further, the human agent will likely have greater social influence, leading to more trust in the human than the software. There might be individual differences based on participants' familiarity with technology, but overall, we expected humans to be perceived more favorably than software. We also expected that more interactive software would be more trustworthy than less interactive software.

Law and Human Oversight

Perhaps to the dismay of some individuals, tax software does not just generate a tax return without the involvement of a human (tax preparer or taxpayer). Thus, it is also important to consider the role of human oversight in perceptions of trust. Specifically in terms of legal decisions (which include tax decisions), research has indicated that software is perceived as more trustworthy and just with human oversight. For example, algorithms have been developed to assess defendant risk and replace judicial bail determinations in some jurisdictions (Kleinberg et al., 2018). However, in practice, researchers have found that the algorithms are plagued by human bias—risk assessment tools rely on past data that were biased and subjective, resulting in continued disproportionate harm to Black and Latinx people (Hill, 2021). Scholars have therefore argued that legal systems should incorporate technology in combination with oversight by human experts, such as judges, to promote fairness (Xu et al., 2021). If we apply this to tax software, it is possible that people would trust software more if a competent human expert oversees its decisions.

Experimental research on citizens' perceptions of alternative dispute resolution methods supports the idea that software can be perceived as beneficial when it is not the ultimate decision maker. Sela (2018) manipulated alternative dispute resolution method (mediation, arbitration) and facilitator (human, software) and measured perceptions of procedural justice. When the facilitator made a decision (arbitration), participants rated the process as more procedurally just with the human than the software; when the facilitator merely guided the process (mediation), participants rated it as more procedurally just with the software than the human (Sela, 2018). Thus, even laypeople have an intuition that computer-assisted legal decisions are fairer, but not if computers are making the ultimate decision.

Previous research has focused on expert oversight, which might create problems for the use of tax software by a taxpayer. At this point, both tax professionals and taxpayers prepare taxes using software, and they e-file at roughly equal rates (see Figure 1). However, the quality is likely vastly different. Tax professionals rely on their expertise to guide the process (e.g., selecting which forms to include), whereas taxpayers might not have sufficient knowledge to be able to provide meaningful oversight. There has been limited research on American taxpayer knowledge—although we can assess self-reported subjective knowledge, there is no existing objective scale of tax knowledge (likely because tax law changes annually). One study measured taxpayer knowledge through assessing the accuracy of tax return estimates (Porto & Collins, 2017). Findings indicated that taxpayers were hesitant to provide an estimate and were often inaccurate in the estimates they did provide. Young taxpayers in the study were more likely to provide estimates of their returns than older taxpayers, but when estimates were compared, older taxpayers' estimates were significantly more accurate than those of younger taxpayers (Porto & Collins, 2017). Thus, self-report and accuracy of return estimates are two possible (admittedly imperfect) ways to assess tax knowledge, which might be important in

considering whether the taxpayer can provide sufficient expertise in order to oversee their taxes.

The Present Study

Whereas existing research has established that trust and procedural justice are crucial to tax compliance, the role of tax software—beyond generic e-filing—remains underexplored. Moreover, much of the prior research has been survey based, nonexperimental, and conducted outside the United States. With this study, we sought to fill this gap by examining how tax software influences taxpayer perceptions and behavior. We focused on four primary research questions: First, how do perceptions of human tax experts compare with those of tax software? Second, do perceptions vary across different types of tax software? Third, do trust and procedural justice predict filing decisions? Fourth, are taxpayers able to effectively oversee tax preparation software? To address these questions, we developed six hypotheses based on prior research.

Hypothesis 1: Perceptions of Humans Versus Software

Prior research indicates that humans are perceived more favorably than software in complex processes (Nagtegaal, 2021) and that more interactive software is perceived more favorably than less interactive software (Blascovich et al., 2002; Nowak & Fox, 2018). Therefore, we hypothesized that participants would have the most favorable perceptions of the human tax preparer, followed by the (more interactive) commercial software and then the (point-and-click) IRS software.

Hypothesis 2: Procedural Justice and Satisfaction

Previous research has indicated that higher perceptions of procedural justice are linked to greater satisfaction with the outcomes (Landis & Goodstein, 1986; Shestowsky, 2014; Tyler, 1984). Therefore, we expected that participants with higher perceptions of procedural justice in the tax system would have higher overall satisfaction.

Hypothesis 3: Knowledge and Expectations

We were interested in the accuracy of taxpayer expectations, on the basis of the research of human oversight (Porto & Collins, 2017). We predicted that participants with greater self-identified knowledge about the tax system would have more accurate expectations about their outcomes than those with lower self-identified knowledge.

Hypotheses 4–6: Filing Behavior

Finally, we were interested in examining decisions to file with software (vs. an expert) or not to file with any method. We had competing predictions about whether process (i.e., procedural justice) or outcome would be more important to participants' selection, if they decide to file (Hypothesis 4). If process matters most, we expected that participants would file with the method they rated as most trustworthy (Hypothesis 4a); if outcome matters most, we expected that participants would file with the method that was most financially beneficial (Hypothesis 4b).

We also expected that participants with higher trust in software would be more likely to file using software than participants with lower trust in software (Hypothesis 5). Finally, given the importance of consistency and trust, we expected that participants would be less likely to file if the outcomes generated by the methods (especially the expert and commercial software) were inconsistent than if they were consistent (Hypothesis 6).

Method

Participants

An a priori power analysis for an F test conducted in the R software environment with the "pwr" package (Version 4.2.2; R Core Team, 2022) indicated that 130 participants were necessary to have 80% power to detect a medium effect size ($f = .25$). We preregistered a sampling plan to recruit up to 155 participants (130% + 20% to account for attrition) during weekend sessions between January 31, 2024 (the legal deadline for release of tax forms), and April 15, 2024 (the legal deadline for tax filing), stopping either when we reached 155 participants or at the end of the time period, whichever came first. The time period ended before we were able to recruit 155 participants. Therefore, participants in the study were 146 taxpayers recruited from the community who participated during the 2023 tax season.

To be considered eligible, participants had to be U.S. citizens or permanent residents who needed to file taxes. To minimize confounds, we recruited only participants who were not self-employed or independent contractors and who did not have any investments (stocks, home, rental property, cryptocurrency, health savings account). Most participants were eligible for the free version of the commercial software on the basis of these exclusion criteria; however, we covered the cost to upgrade to the paid version when necessary ($n = 15$). Participants also had to participate in person, so we limited participation to taxpayers local to the research project. Thus, all participants were residents of Texas, which has no state income tax, so we were able to focus exclusively on the federal filing. We included data from all 146 taxpayers in this study, although some had missing data or forms and were therefore excluded from specific analyses.

Taxpayers in this study were mostly women (64%; 36% men), Hispanic (88%; 12% non-Hispanic), and White (75%; 1.4% Black, 1.4% American Indian/Alaska Native, 0.7% Asian, 9% other, 10% prefer not to say) and reported being neither conservative nor liberal (25%; 32% slightly to moderately liberal, 10% slightly to moderately conservative, 18% extremely liberal, 12% prefer not to say). The average age of participants was 27.55 years ($SD = 9.09$; range = 18–64), and most were not married (89%) and did not have kids (88%; 4.8% one child, 5.5% two children, 1.4% three children). Most participants (92%) were considered to be of low socioeconomic status (SES; income: M = \$25,573.75, SD = \$34,440.66, Mdn = \$19,529.56, range = \$0–\$328,999). Although we were not specifically recruiting students, 55% of participants were students (43% full time; 12% part time).

Materials

All materials are provided on the Open Science Framework (https://osf.io/emrxs). Here, we describe the materials that are of

interest in this article. The institutional review board at the University of Texas at El Paso reviewed this study, which was deemed exempt (institutional review board approval 2074783).

Initial Expectations

Prior to completing the study, participants answered questions about their initial expectations. This included how they completed their taxes last year, whether their situation changed between last year and this year, what they expected their return amount to be, and which method they expected to prefer out of the three they would be using. We also asked them to rate their knowledge of the tax preparation process on a 5-point Likert-type scale ranging from *far below average* (1) to *far above average* (5), with a midpoint of *average* (3).

Tax Preparation Method and Questions

Participants prepared their taxes in three different ways in a randomly assigned order: a human tax preparer, popular commercial software, and free IRS software. After completing each tax preparation method, participants rated the method in trustworthiness, satisfaction, and confusion on a 5-point Likert-type scale ranging from *not at all* (1) to *extremely* (5).

For the human, we hired a local tax service that provided the tax preparation services of two licensed tax experts on the weekend for a fee paid by us (M = \$202.06, range = \$35–\$373).

For the popular commercial software, we selected one of the top three tax preparation services. Most participants were eligible for the free version, but we paid for upgrades for 15 participants (range = \$39–\$79). Two participants were ineligible for the study but did not disclose it in the eligibility questionnaire, and the other 13 participants were upgraded because of additional forms that were not considered in the eligibility questionnaire (eight premium tax credits, Form 8962; four qualified retirement savings beyond standard retirement, Form 8880; one alternate minimum tax, Schedule 2).

For the IRS software, we used the IRS Free Fillable Forms. This is a free option available to all taxpayers (https://www.freefilefillableforms.com/home/default.php).

Filing and Ranking Decisions

After completing all three methods, participants were asked which method they would prefer to file with, if at all. If they indicated that they would not use any of these methods to file, they were asked how they would prefer to file and which method they would pick if they had to select one of the three methods. Participants then ranked the three methods in five categories: trustworthiness, satisfaction, ease of use, confusion, and intent to use in the future. For the rankings, 1 indicated the method they believed fit the category the best and 3 indicated the method they believed fit the category the worst.

Perceptions and Demographics

Participants next completed three scales that assessed perceptions of taxes and software, presented in random order. Participants then provided standard demographic information (age, gender, race, ethnicity, number of children, political affiliation).

Procedural Justice and Taxes. To assess perceptions of procedural justice in tax preparation, we had participants complete a 10-item Tax Authority Procedural Justice Scale (Murphy et al., 2016). Participants rated statements (e.g., "The IRS tries to be fair when making decisions") on a 5-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (5)—all points on the scale were labeled, with a midpoint of *neither agree nor disagree*—and ratings were then averaged to create a total score. Higher scores indicated higher perceptions of procedural justice within the tax system (M = 2.76, SD = 0.45, range = 1–5). The scale has been found to be reliable (α = .88) and was reliable in our sample (α = .90).

e-File Trustworthiness. To assess perceptions of trustworthiness of e-filing, we had participants complete the 24-item Trust in e-Filing Scale (Carter et al., 2011). Participants rated statements (e.g., "I find an e-file system easy to use") on a 5-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (5)—all points on the scale were labeled, with a midpoint of *neither agree nor disagree*—and ratings were then averaged to create a total score. Higher scores indicated more favorable perceptions of e-filing (M = 3.38, SD = 0.35, range = 1–5). This scale was reliable in our sample (α = .94).

Computer Trustworthiness. To assess trust in tax software, we had participants complete the 12-item Human–Computer Trust Scale (Gulati et al., 2019). This scale is designed to be modified to directly address the relevant interaction, so we focused on the use of tax software. Participants rated statements (e.g., "I can trust the information presented to me by tax software") on a 5-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (5)—all points on the scale were labeled, with a midpoint of *neither agree nor disagree*—and ratings were then averaged to create a total score. Higher scores indicated more trustworthiness in interactions with tax software (M = 2.89, SD = 0.45, range = 1–5). This scale was reliable in our sample (α = .91).

Design

The present study followed a within-subjects design with three levels of tax preparation method (human tax expert, commercial software, IRS software).

Procedure

Participants were recruited using flyers, social media (Instagram, Facebook, Reddit, Craigslist, and NextDoor), and email. The recruitment materials directed them to an eligibility questionnaire. Eligible participants were then contacted to schedule a session in person at the university on weekends.

When they arrived, participants provided informed consent and approved the use of their data. Participants then answered questions about their initial expectations. Then, they completed their taxes in three different ways in a randomly assigned order and completed the tax preparation method questionnaire after each. Research assistants monitored participants while they were using each method to ensure that there were no technical difficulties, but we emphasized to participants that we were unable to provide legal tax advice or answer any of their tax-related questions. They were free to ask the independent tax preparer questions during their session or after the study. After completing all three methods, participants completed a

final questionnaire with their filing and ranking decisions, scales, and demographic information. Finally, we debriefed participants and paid them. Participants did not file their taxes during the study, but they were given the opportunity to file their own taxes after the study was completed if they wanted.

Results

Before testing our hypotheses, we present descriptive statistics for context. Most participants (55%) said that their situation had not changed since last year, but slightly more than one quarter (28%) said that it had changed, and 13% were unsure of whether it had changed. Table 1 presents descriptive statistics in relation to each method. We paid for all preparation methods in this study, which cost an average of $202.26 for the tax preparer and $5.20 for the commercial software ($0 for IRS software; see Table 1). Most of our participants were able to prepare their taxes for free with the commercial software, because of our selection criteria. It is difficult to determine the value of the professional, but when looking at average net return after including costs, we found that participants netted significantly more using the commercial software (M = $499.72, SD = $834.34) than the human tax preparer (M = $351.25, SD = $822.86), $t(87) = 2.23$, $p = .03$, Cohen's $d = 0.24$, 95% confidence interval [CI: 0.03, 0.45]. We also conducted all analyses controlling for various demographics, but the results were not substantially different, and therefore, these results are available at Open Science Framework (https://osf.io/emrxs/) (controlling for political ideology: addtional online material 1; comparing students with nonstudents: additional online material 2; controlling for income: additional online material 3; comparing tax situation change with no change: additional online material 4; excluding upgrades: additional online material 5).

Hypothesis 1: Perceptions of Humans Versus Software

First, we tested whether participants had the most favorable perceptions of the human tax preparer, followed by the (more interactive) commercial software and then the (point-and-click) IRS software. In our preregistration, we defined more favorable perceptions as higher trust, higher satisfaction, greater ease, and less confusion. We conducted a series of rank-order analyses using Friedman's tests and paired-samples t tests (see Table 2). As expected, across all variables, the human tax preparer was rated most favorably, followed by the commercial software and then the IRS software. It is worth noting that, at the beginning of the study, 21% of participants reported expecting to like the IRS software the best (see Table 1). We also explored intent to use in the future—IRS software was ranked the lowest, but there was no significant difference between the commercial software and the human tax preparer.

Hypothesis 2: Procedural Justice and Satisfaction

Next, we tested whether participants with higher perceptions of procedural justice in the tax system would have higher overall satisfaction. A correlation between scores on the Tax Authority Procedural Justice Scale and satisfaction scores for the method that participants opted to use (excluding those who did not want to file) did not support this hypothesis. Contrary to expectations, the correlation was not significant, $r(131) = .12$, $p = .17$, 95% CI [−0.05, 0.29].

Hypothesis 3: Knowledge and Expectations

To test the relationship between self-identified knowledge and expected return accuracy, we calculated an expectation accuracy score for each method by taking the absolute value of the difference between their estimate of the return and the actual return amount of the method. Participants self-reported that knowledge of the tax preparation process was below the mean (M = 2.84, SD = 0.91, range = 1–5), and only 59 participants (46%) estimated their return amount (M = $716, SD = $1,514). There were also extreme outliers, so we excluded data 3 SDs above or below the mean. On average, estimations were off by $510.85 ($SD$ = $793.44, $n_{returns}$ = 51) compared with the expert, $577.85 ($SD$ = $873.57, $n_{returns}$ = 54) compared with the commercial software, and $510.85 ($SD$ = $1,436.74, $n_{returns}$ = 55) compared with the IRS software.

Contrary to expectations, there was no relationship between self-identified knowledge about the tax system and expectation accuracy. None of the correlations between self-identified knowledge and

Table 1
Descriptive Information for Each Filing Method

Outcome measure	Human tax preparer	Commercial software	IRS software
Used last year	27%	33%	0%
Expect to prefer	48%	31%	21%
Used to file this year	71%	20%	2%
Median return amount	$333.00	$252.50	$220.50
	[−$2,781 to $10,125]	[−$1,552,461 to $10,125]	[−$28,168 to $23,056]
Mean cost of preparation	$202.26 ($52.83)	$4.62 ($14.39)	$0
	[$35–$373]	[$0–$89]	[$0–$0]
Mean time to prepare	25.51 (11.94)	30.10 (14.12)	31.31 (24.54)
	[9–80]	[5–93]	[4–172]
Mean number of forms	1.68 (0.79)	1.97 (1.73)	1.10 (0.35)
	[1–4]	[1–9]	[1–3]

Note. Means are reported with standard deviations in parentheses and ranges in brackets. All times are in minutes. For commercial software, only 15 returns were upgraded from the free account; eligible participants should have been able to use the free version but were not excluded if they were upgraded after they started the study. Percentages not summing to 100% in the prior year filing were due to participants not filing or reporting another method. Commercial software used in prior years includes all, not only the one used in the study. IRS = Internal Revenue Service.

Table 2
Hypothesis 1—Rank Order Analysis of Preparation Methods

	Human	Commercial	IRS	Human versus commercial	Human versus IRS	Commercial versus IRS
Outcome measure	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*t* (*p*)	*t* (*p*)	*t* (*p*)
Trustworthiness	1.13 (0.39)	2.05 (0.50)	2.78 (0.55)	15.69 (<.001)	25.64 (<.001)	9.80 (<.001)
Satisfaction	1.20 (0.44)	1.91 (0.56)	2.84 (0.47)	9.56 (<.001)	29.83 (<.001)	12.90 (<.001)
Ease of use	1.22 (0.46)	1.85 (0.51)	2.90 (0.41)	8.58 (<.001)	31.62 (<.001)	18.04 (<.001)
Confusion	2.75 (0.55)	2.09 (0.48)	1.11 (0.43)	9.38 (<.001)	23.86 (<.001)	16.45 (<.001)
Future use	1.54 (0.63)	1.65 (0.67)	2.75 (0.60)	1.00 (.16)	13.26 (<.001)	11.19 (<.001)

Note. Ranks are coded such that 1 is highest and 3 is lowest. Friedman tests indicated that all rankings were significantly different prior to paired-samples *t*-test pairwise comparisons—trustworthiness: $\chi^2(2, 141) = 194.76$, $p < .001$; satisfaction: $\chi^2(2, 139) = 189.44$, $p < .001$; ease: $\chi^2(2, 143) = 207.04$, $p < .001$; confusion: $\chi^2(2, 142) = 194.91$, $p < .001$; future use: $\chi^2(2, 109) = 98.67$, $p < .001$. IRS = Internal Revenue Service.

accuracy scores were significant—expert: $r(52) = .26$, $p = .06$, 95% CI [−0.02, 0.50]; commercial software: $r(53) = .24$, $p = .09$, 95% CI [−0.04, 0.48]; IRS software: $r(66) = -.05$, $p = .70$, 95% CI [−0.29, 0.20].

Hypotheses 4–6: Filing Behavior

Finally, we examined tax filing behavior through a series of binary logistic regressions. One hundred thirty-three participants (94%) reported wanting to file with one of these methods. Of participants who decided to file, 76% filed with the expert, 22% filed with the commercial software, and 2.3% filed using the IRS software.

To test Hypothesis 4, we conducted analyses with tax filing method (human or commercial software) as the outcome and trustworthiness or outcome of each method as the predictors; because so few people opted to file with the IRS software, participants who chose the IRS software or opted not to file were excluded. To test Hypothesis 5, we conducted the analysis with software filing (no or yes) as the outcome (collapsing human filing or no filing into "no" and commercial and IRS form software into "yes") and the computer perception scales (e-file trust and computer trust) as the predictors. To test Hypothesis 6, we conducted the analysis with filing decision (no or yes) as the outcome (collapsing expert and software into "yes") and the absolute value of the difference between the return amount generated by the human and that generated by the commercial software as the predictor. The three participants who chose to file with the IRS software were excluded because of low group size and inconsistent return amounts.

Trust was more important to filing decision than outcome, consistent with the procedural justice hypothesis (Hypothesis 4a). However, only trustworthiness of the software mattered (see Table 3). Participants with higher trust in the commercial software had more than three times the odds of filing with the commercial software over the human expert; however, there was not a similar relationship between trust in the human expert and filing with the human expert.

Participants with more trust in e-filing were 2.61 times more likely to file using the commercial software than the human expert, $\beta(1) = 0.96$, $SE = 0.39$, $p = .01$, $\exp(\beta) = 2.61$, 95% CI [1.22, 5.58], which was partially consistent with expectations (Hypothesis 5). However, the human–computer trustworthiness scale did not significantly predict filing choice, $\beta(1) = 0.44$, $SE = 0.32$, $p = .17$, $\exp(\beta) = 1.55$, 95% CI [0.83, 2.89].

Finally, outcome consistency did not predict whether the participant chose to file their returns using one of the three options, $\beta(1) = 0.00$, $SE = 0.00$, $p = .87$, $\exp(\beta) = 1.00$, 95% CI [1.00, 1.00], which was contrary to expectations (Hypothesis 6). Overall, fewer than half of participants had return amounts that differed between the two methods (43%), and a large majority of participants chose to file with one of the two methods (98%).

Discussion

The present study is the first to experimentally compare tax preparation methods by American taxpayers with a focus on comparing experts and software. We had four research questions guiding this project: (1) How do perceptions of human tax experts compare

Table 3
Hypothesis 4—Prediction of Filing Choice Comparing Trust and Outcomes

Model variable	β	*SE*	*df*	*p*	exp(β)	95% CI
Model 1—trust						
Trust of human expert	−0.74	0.42	1	.081	0.48	[0.21, 1.10]
Trust of commercial software	**1.20**	**0.29**	**1**	**<.001**	**3.32**	**[1.88, 5.86]**
Model 2—outcomes						
Return from human expert	0.00	0.00	1	.20	1.00	[1.00, 1.00]
Return from commercial software	0.00	0.00	1	.78	1.00	[1.00, 1.00]

Note. Bolded lines indicate significant effects ($p < .05$). Model 1: goodness of fit $\chi^2(2) = 24.19$, $p < .001$; −2 log likelihood = 112.28; Cox and Snell $R^2 = .17$; Nagelkerke $R^2 = .26$. Model 2: goodness of fit $\chi^2(2) = 2.66$, $p = .26$; −2 log likelihood = 116.00; Cox and Snell $R^2 = .03$; Nagelkerke $R^2 = .04$. *SE* = standard error; exp(β) = odds ratio; CI = confidence interval.

with tax software? (2) Do perceptions of different types of tax software differ? (3) Do trust and procedural justice predict filing decisions? and (4) Can taxpayers sufficiently oversee tax preparation software?

Tax Experts Are Generally Preferred Over Tax Software

The first question we addressed in this study was how professional tax preparers were perceived compared with tax software. As expected (Hypothesis 1), participants generally ranked the expert the most favorable, followed by the commercial software and then the IRS software. Participants were also most likely to file with the expert, followed by the commercial software and then the IRS software. Future research could examine how much of the increased ratings are due to questions being answered during the preparation process. Although we did not examine it directly in the study, during the debrief, many participants indicated to us that they are generally very confused by taxes and find that the software can answer only a limited number of questions they have. Many participants told us that they appreciated interacting with the tax expert because the tax expert explained different terms and answered their questions. Many said that they felt that the guidance they got from the expert during this study would help them in their tax preparation in the future. In fact, one participant who was receiving a large refund but struggled paying bills throughout the year told us that the expert gave them guidance on what to discuss with their employer to change the withholding so they would have more money throughout the year—advice they deemed life changing.

The only category in which experts were not rated significantly better than the commercial software was the method they intended to file with next year. Participants were equally likely to indicate that they would file with the expert as with the commercial software in the future. One factor that might be playing a role in this decision is the cost of preparation. We paid for the cost of tax preparation this year, but in future years, the taxpayer will be responsible for their own preparation costs (despite many participants requesting to participate in this project again next year). Thus, the cost of hiring a professional might be a barrier for many of these taxpayers in the future. Anecdotally, several taxpayers noted that they learned so much from the professional that it will assist them in preparing their own taxes in the future, even using software. Thus, future research should examine how workshops with tax preparers could improve long-term outcomes for taxpayers.

Interactive Software Is More Preferable Than Point-and-Click Software

The second question we addressed in this project was how different types of tax software were perceived. As expected, participants had more favorable perceptions of the interactive commercial software compared with the point-and-click IRS software (Hypothesis 1). Participants were also far more likely to file with the commercial software than the IRS software. Whereas the commercial software asks taxpayers questions and walks them through each step, the IRS form is more difficult to use and requires the taxpayer to know exactly what information to provide. The associated handbook to answer questions is several hundred pages long and complex to navigate. Anecdotally, the most common complaint from participants during our debriefing sessions was how difficult it was to use the IRS software.

We used only one type of commercial tax software in this study to maintain experimental control, but prior research and our experience with participants in this study suggest that it would be worthwhile to examine different types of commercial software as well. Nine participants (6%) decided not to file using any of the methods. Although we are not sure why several individuals made this decision, three told us during debrief that it was because they had strong loyalty to a different commercial software not used in this study, and therefore, they intended to complete their taxes for a fourth time using the other method. It is a small number, but it would be interesting to examine what software people choose when given a choice and whether there are differences in perceptions and outcomes across types of commercial software. Alternatively, examining open-source software that is not owned by the IRS might be good for future research as well (Tizpaz-Niari et al., 2023).

Although the commercial software was much more user friendly than the IRS software, the IRS software should not be completely ruled out. There have been criticisms that commercial software companies trick consumers and lobby the government for financial gain (Elliot & Kiel, 2019). As part of the preparation with the commercial software, each participant had to click through at least four screens asking them to pay to upgrade from the free software. Thus, the increased reliance on commercial software might be detrimental for taxpayers in the long run.

In fact, the IRS recently tested a Free File program (not to be confused with Free Fillable Forms used here) that provides select taxpayers with more guidance in their tax preparation (IRS, 2024b). This new method had limited eligibility and was available only to qualified taxpayers, and its use was not sufficiently widespread to warrant inclusion in our study (we could not guarantee when designing the study how many participants would be eligible). This new option with increased guidance and interactivity might be able to compete with the commercial software—tax preparation companies purportedly spent more than $90 million lobbying against this software (Massoglia, 2023). If the software is comparable with commercial software, it could be particularly useful for people who do not qualify for the free version of the commercial software. Taxpayers might be willing to use IRS options, especially because 21% of participants initially expected to prefer the IRS software (but only 2% ultimately filed with that method). However, more research is needed on taxpayer perceptions of this new technology.

Another alternative that might make IRS software more user friendly is prepopulating tax forms. Here, taxpayers did not like the point-and-click IRS software that required them to prepare an entire tax return with little guidance. However, many other wealthy countries use prepopulated tax forms that require taxpayers to simply correct the information provided by the country's tax authority, rather than having the taxpayer complete the return from the start on their own (Elliot & Kiel, 2019). Point-and-click IRS software might be more favorable if used only to correct prepopulated returns rather than creating the return from scratch. Moreover, research on prepopulating tax returns has found that not only is the process of checking the return easier for taxpayers than preparing the return on their own, but it also leads to greater accuracy (van Dijk et al., 2020). Participants were more attentive to inaccuracies on prepopulated forms and tax compliance improved (van Dijk et al., 2020). The commercial tax software lobby has blocked such procedures in the United States to protect their financial

interests (Elliot & Kiel, 2019), but if such a procedure were adopted, taxpayer perceptions of the IRS software might change.

Perceptions of Trust and Procedural Justice Affect Tax Preparation Decisions

Our third question was how trust and procedural justice related to taxpayer perceptions and behaviors. In terms of filing decision, the procedural justice hypothesis was supported (Hypothesis 4a)—participants' filing decisions were more influenced by trust than outcome. In fact, the outcome did not significantly predict filing decision at all. However, it appears that trust in software was the driving factor; trust in the expert did not predict filing decision. Both trust in the specific commercial software used in this study (Hypothesis 4) and general trust in software (Hypothesis 5) predicted participants' decisions to file using the software. Given the high rates of filing using the tax expert, it appears that using the expert was the default that was overcome only by high trust in software.

However, contrary to expectations, perceptions of procedural justice of the tax system did not significantly correlate with satisfaction in this study (Hypothesis 2). The lack of support for this hypothesis might be due to some differences in measurement compared with other studies. First, we assessed satisfaction rather than compliance, which was the focus of prior tax research. Although prior research on criminal defendants still supports the link between procedural justice and satisfaction (Tyler, 1984), it is possible that tax law is different. People who have higher procedural justice scores might be more likely to comply with the law but still might not be fully satisfied that they have to pay taxes (or are getting a smaller refund than desired). We did not ask participants whether they intended to comply with tax law because of ethical concerns.

Second, future research could examine different measures of satisfaction. In this study, our focus was on taxpayers' preference of methods, so our satisfaction measure reflected this. In the future, researchers could also examine more general satisfaction. Rather than including only assessments of the specific preparation methods, researchers could also assess perceptions of and satisfaction with the IRS more broadly.

Inconsistencies across methods did not predict participants opting not to file using any of the methods. Most participants chose to file using one of the study methods, and outcomes were generally consistent between the software and expert. In the future, research should examine this question with more complicated tax returns that have the potential to have more inconsistencies.

Taxpayer Oversight Warrants Further Study

The final question in this research project was whether taxpayers can provide sufficient oversight of tax software to eliminate concerns. We did not find support for our hypothesis that there would be a correlation between self-reported knowledge and expectations (Hypothesis 3). This was a novel question in this area and was subject to several limitations.

First, most people declined to estimate their expected return. Although this is consistent with prior research (Porto & Collins, 2017), our analyses were restricted to less than half our sample. Thus, the analysis was underpowered. In the future, researchers could do more to encourage participants to provide estimates, but forcing responses is ethically questionable and it might be important to control for whether participants were willing to provide an estimate without pressure. Moreover, the mere fact that many participants were hesitant to provide an estimate might reflect low knowledge of the tax process and inability to oversee the software. If taxpayers do not have any expectations about how much they should be getting, it might be difficult to determine whether the software is providing an accurate return.

Second, we used a self-report measure of tax knowledge with mean scores that were below the midpoint. Thus, participants are reporting not knowing much about their taxes. Although subjective self-report scores can correlate with objective knowledge (e.g., numeracy; Liberali et al., 2012), generally the correlations are not perfect. For example, self-report numeracy scores are often prone to self-reporting biases such as the Dunning-Krueger overestimation-of-ability effect (Reyna et al., 2009). With taxes, it might be that people are overestimating their ability and are actually much worse or that they are underestimating their abilities. However, it is also extremely difficult to objectively measure tax knowledge. Tax law is complicated—it takes professionals years of training to become experts in the subject—and many rules change annually. Also, it is arguable whether the average taxpayer needs to know everything about tax law; knowing the rules that apply to them is often sufficient. Nevertheless, the low average self-report tax knowledge scores are likely concerning for a taxpayer's ability to oversee tax software. It appears that, in this study, many participants did not have sufficient tax knowledge to confidently oversee their tax preparation.

Finally, it is also important to note that in order to maintain experimental control, this study limited the participants to those with simple tax returns. As tax returns become more complicated, knowledge might become even more important. On average, our participants had fewer than two tax return forms; taxpayers with more complicated taxes have significantly more forms to oversee and ensure that they are entered accurately. More complicated tax returns also likely take longer than the 30 min it took on average for participants to prepare their taxes using the software in this study. Including a wider variety of taxpayers in future research would be beneficial, especially for this question.

Although this correlation was not significant, future research could also examine whether increasing taxpayer knowledge assists with the process. Although tax rules change annually, and each taxpayer has a different situation, it might be possible to develop a training that focuses on common tax situations that tend to be legally consistent over time. Future research could test the long-term benefit of such a training or workshop to see whether it improves tax preparation, particularly for people of low SES backgrounds or first-time tax filers. Some of our participants anecdotally suggested that the interaction with the tax preparer would be helpful for them preparing their taxes in the future. Unfortunately, we do not have the ability to conduct a follow-up study with participants next year to see whether the one-time interaction with the tax experts helped participants better understand their taxes long term, but future research could experimentally examine this question.

Future Research

This study brings up several questions, in addition to the future research discussed previously. For example, one of the strengths of this study is that it primarily used participants of low SES. Psychological research has been criticized for focusing on a narrow

range of people, and one group that is often overlooked is people of low SES (Emery et al., 2023). Given that people of low SES make up 30% of the 135.6 million people responsible for taxes every year, this is an important group to study, particularly in the tax domain. Financial decisions have higher stakes for people from lower SES backgrounds (linked to mental health issues, suicide, and homelessness; Mullainathan & Shafir, 2010). The cost of preparing taxes is proportionally higher for people with fewer financial resources, and they are consequently more likely to use software than pay an accountant. Moreover, this group is particularly at risk of an audit, especially since funding to the IRS decreased (Rampell, 2024). Although, as we noted earlier, this might result in limits to generalizability (just to a different group from most psychological research) and we suggest that future research examine other tax brackets, we also think that future research should focus more on this group of taxpayers. Specifically, this study suggests that people from lower income groups might need more accessible free tax preparation methods. Therefore, future research could test how programs such as the IRS Volunteer Income Tax Assistance program, which provides communities with resources to assist people with low income in preparing their taxes, compare with the free software.

Another direction for future research is examining accuracy of the tax returns in addition to taxpayer perceptions and filing behavior. Although for most returns, the tax expert and commercial software generated similar numbers, there were still some differences. Therefore, future research would benefit from better testing accuracy of the different methods. It might even be possible to develop test cases that include common errors as identified by the IRS. Even if software were not directly compared with humans, it would be beneficial to have a better understanding of where the software errors exist. Although companies might do this internally, it would be beneficial for consumers to know as well. For example, in the past, we have tested open-source software using metamorphic testing methods that allowed us to identify areas in which the software provided inconsistent outcomes (Tizpaz-Niari et al., 2023). Using similar methods with private commercial software could be beneficial.

Future research could examine how software design might influence accuracy and compliance with tax returns. Most commercial software uses a method that shows the taxpayer the return amount at the top of the screen and adjusts the amount as the taxpayer enters information. This design might encourage taxpayers to alter the way they enter information if the adjustment is unfavorable. Especially because our research indicates that people have low knowledge and are hesitant to provide expectations, it is possible that people do not know what the answer should be, and therefore, they enter information and adjust it until they obtain the most beneficial outcome.

Another variable that could be interesting in future research is taxpayer perceptions of the likelihood of being audited. Future research could explore the connection between perceptions of likelihood of being audited and decisions to file using a particular method. For example, if different tax preparation methods result in different outcomes, then it is possible that the decision to file would be based on not just which method is preferred or generates the best outcome but which option has the least risk. This might be particularly interesting to examine in combination with general risk-seeking behavior.

Finally, future research should examine how artificial intelligence and machine learning alter the tax preparation process. The present study looked only at rule-based software (software that follows a fixed set of patterns to derive outcomes), but an emerging body of literature considers trust and procedural justice with more advanced models such as deep neural networks and large language models that may better mirror human experts because of their data-driven nature (Acikgoz et al., 2020; De Cremer et al., 2024; Liu, 2021). These newer methods might increase user satisfaction because they are more interactive, can answer more user questions, predict more taxpayer circumstances, and might even give taxpayers more control. But they might not be more accurate than traditional software and might be more difficult to audit and test because of their uninterpretable, black box nature (Rudin, 2019). Therefore, it will be extremely important to ensure that machine learning and artificial intelligence methods are accountable.

Conclusion

The increasing reliance on tax software among taxpayers necessitates a deeper understanding of how these tools are perceived and trusted. This study reveals that although human tax experts are generally preferred, there remains significant trust in commercial tax software, with many taxpayers expressing a willingness to continue using it in the future, likely because of the cost effectiveness compared with professional services.

As tax software becomes more integral to the filing process, particularly among low-income individuals, it is crucial to ensure that these systems are accountable. The responsibility for any errors or bugs ultimately rests with the taxpayer (*Langley v. Comm'r*, 2013), placing a heavy burden on individuals who may not fully understand the software they rely on. Although tax professionals often share legal responsibility for the returns they prepare, commercial software generally offers protection only at an additional cost, and IRS software leaves the taxpayer fully accountable.

Building on Alexander Hamilton's assertion that taxation is an indispensable ingredient of our Constitution, it is evident that tax compliance is fundamental to the stability and security of the nation. In this digital age, tax software plays a critical role in this process, serving as the modern bridge between citizens and the financial obligations that support our government. Therefore, it is imperative that tax software not only facilitates compliance but also upholds the principles of trust, fairness, and justice upon which the Constitution is founded. Ensuring that these tools are reliable and just is not merely a matter of convenience; it is a responsibility to safeguard the very foundations of our democratic system.

References

Abric, J. C., & Kahan, J. P. (1972). The effects of representations and behavior in experimental games. *European Journal of Social Psychology*, *2*(2), 129–144. https://doi.org/10.1002/ejsp.2420020203

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Akram, M. S., Malik, A., Shareef, M. A., & Goraya, M. A. S. (2019). Exploring the interrelationships between technological predictors and behavioral mediators in online tax filing: The moderating role of perceived

risk. *Government Information Quarterly*, *36*(2), 237–251. https://doi.org/10.1016/j.giq.2018.12.007

Alm, J., & Torgler, B. (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology*, *27*(2), 224–246. https://doi.org/10.1016/j.joep.2005.09.002

Araujo, T., Brosius, A., Goldberg, A. C., Möller, J., & de Vreese, C. (2023). Humans vs. AI: The role of trust, political attitudes, and individual characteristics on perceptions about automated decision making across Europe. *International Journal of Communication*, *17*, 6222–6249. https://ijoc.org/index.php/ijoc/article/view/20612/4348

Attempt to evade or defeat tax, 26 U.S.C. § 7201. (1954). https://www.law.cornell.edu/uscode/text/26/7201

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Carter, L., Christian Shaupp, L., Hobbs, J., & Campbell, R. (2011). The role of security and trust in the adoption of online tax filing. *Transforming Government*, *5*(4), 303–318. https://doi.org/10.1108/17506161111173568

Chaouali, W., Ben Yahia, I., & Souiden, N. (2016). The interplay of counter-conformity motivation, social influence, and trust in customers' intention to adopt Internet banking services: The case of an emerging country. *Journal of Retailing and Consumer Services*, *28*, 209–218. https://doi.org/10.1016/j.jretconser.2015.10.007

Choung, H., Seberger, J. S., & David, P. (2024). When AI is perceived to be fairer than a human: Understanding perceptions of algorithmic decisions in a job application context. *International Journal of Human–Computer Interaction*, *40*(22), 7451–7468. https://doi.org/10.1080/10447318.2023.2266244

Chu, X., Haynie, J. J., & Zhang, Y. (2022). Information collection concern, procedural justice, and intention to use online tax software: A process model. *Journal of Accounting and Taxation*, *14*(4), 304–311. https://academicjournals.org/journal/JAT/article-abstract/860B02969797

Collins, J. (2024, March 20). More taxpayers opt for free tax software this season. *Civic Science*. Retrieved October 28, 2024, from https://civicscience.com/more-taxpayers-opt-for-free-tax-software-this-season/#:~:text=Intuit%20TurboTax%20is%20the%20most,Cash%20%20App%20Taxes%20(16%25)

D'Attoma, J. (2020). More bang for your buck: Tax compliance in the United States and Italy. *Journal of Public Policy*, *40*(1), 1–24. https://doi.org/10.1017/S0143814X18000302

De Cremer, D., Narayanan, D., Nagpal, M., McGuire, J., & Schweitzer, S. (2024). AI fairness in action: A human–computer perspective on AI fairness in organizations and society. *International Journal of Human–Computer Interaction*, *40*(1), 1–3. https://doi.org/10.1080/10447318.2023.2273673

Elliot, J., & Kiel, P. (2019). *Inside TurboTax's 20-year fight to stop Americans from filing their taxes for free*. ProPublica. https://www.propublica.org/article/inside-turbotax-20-year-fight-to-stop-americans-from-filing-their-taxes-for-free

ElSayad, G., & Mamdouh, H. (2024). Are young adult consumers ready to be intelligent shoppers? The importance of perceived trust and the usefulness of AI-powered retail platforms in shaping purchase intention. *Young Consumers*, *25*(6), 969–989. https://doi.org/10.1108/YC-02-2024-1991

Emery, L. F., Silverman, D. M., & Carey, R. M. (2023). Conducting research with people in lower-socioeconomic-status contexts. *Advances in Methods and Practices in Psychological Science*, *6*(4). https://doi.org/10.1177/25152459231193044

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction*, *30*(5), 401–432. https://doi.org/10.1080/07370024.2014.921494

Gangl, K., van Dijk, W. W., van Dijk, E., & Hofmann, E. (2020). Building versus maintaining a perceived confidence-based tax climate: Experimental evidence. *Journal of Economic Psychology*, *81*, Article 102310. https://doi.org/10.1016/j.joep.2020.102310

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human–computer trust scale. *Behaviour & Information Technology*, *38*(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779

Hamilton, A. (1787). *Concerning the general power of taxation* (The Federalist Papers: 30). Lillian Goldman Law Library. Retrieved October 28, 2024, from https://avalon.law.yale.edu/18th_century/fed30.asp

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Report*, *39*, Article 105456. https://doi.org/10.1016/j.clsr.2020.105456

Hill, S. A., II. (2021). Bail reform and the (false) racial promise of algorithmic risk assessment. *UCLA Law Review*, *68*(4), 910–987. https://ssrn.com/abstract=3683790

Ingrams, A., Kaufmann, W., & Jacobs, D. (2022). In AI we trust? Citizen perceptions of AI in government decision making. *Policy and Internet*, *14*(2), 390–409. https://doi.org/10.1002/poi3.276

Internal Revenue Service. (n.d.). *Understanding taxes teacher*. Retrieved October 28, 2024, from https://apps.irs.gov/app/understandingTaxes/teacher/whys_thm01_les01.jsp#:~:text=Taxes%20provide%20revenue%20for%20federal,services%20very%20effectively%20for%20themselves

Internal Revenue Service. (2024a). *Filing season statistics by year*. Retrieved October 28, 2024, from https://www.irs.gov/newsroom/filing-season-statistics-by-year

Internal Revenue Service. (2024b). *Free file: About the free file alliance*. Retrieved October 28, 2024, from https://www.irs.gov/e-file-providers/about-the-free-file-alliance

Internal Revenue Service. (2024c). *IRS annual data book, Table 32*. Retrieved October 28, 2024, from https://www.irs.gov/statistics/soi-tax-stats-irs-data-book-index-of-tables

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

Landis, J. M., & Goodstein, L. (1986). When is justice fair? An integrated approach to the outcome versus procedure debate. *American Bar Foundation Research Journal*, *11*(4), 675–707. https://doi.org/10.1111/j.1747-4469.1986.tb00262.x

Langley v. Comm'r, T.C. Memo. 2013-22 (U.S.T.C. 2013, January 17). https://casetext.com/case/langley-v-commr-1

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgement. *Journal of Behavioral Decision Making*, *25*(4), 361–381. https://doi.org/10.1002/bdm.752

Liu, B. (2021). In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of Computer-Mediated Communication*, *26*(6), 384–402. https://doi.org/10.1093/jcmc/zmab013

Massoglia, A. (2023, September 1). Tax prep companies that spent over $90 million lobbying against free tax-filing system face new scrutiny from lawmakers. *Open Secrets*. Retrieved October 28, 2024, from https://www.opensecrets.org/news/2023/09/tax-prep-companies-lobbying-against-free-file-face-scrutiny-from-lawmakers/

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), Article 12. https://doi.org/10.1145/1985347.1985353

McLeod, A., Pippin, S., & Mason, R. (2009). Individual taxpayer intention to use tax preparation software: Examining experience, trust, and perceived risk. *Journal of Information Science and Technology*, *6*, 1–16. https://www.researchgate.net/publication/266449898_Individual_Taxpayer_Intention_to_Use_Tax_Preparation_Software_Examining_Experience_Trust_and_Perceived_Risk

Mullainathan, S., & Shafir, E. (2010). *Scarcity*. Times Books.

Murphy, K., Bradford, B., & Jackson, J. (2016). Motivating compliance behavior among offenders: Procedural justice or deterrence? *Criminal Justice and Behavior*, *43*(1), 102–118. https://doi.org/10.1177/0093854815611166

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, *38*(1), Article 101536. https://doi.org/10.1016/j.giq.2020.101536

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30–53. https://doi.org/10.12840/issn.2255-4165.2018.06.01.015

Nurkholis, N., Dularif, M., & Rustiarini, N. W. (2020). Tax evasion and service-trust paradigm: A meta-analysis. *Cogent Business & Management*, *7*(1), Article 1827699. https://doi.org/10.1080/23311975.2020.1827699

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in human-machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior*, *89*, 27–39. https://doi.org/10.1016/j.chb.2018.07.022

Picchi, A. (2023, April 13). What happens if you don't file your taxes? The penalties are "huge." *CBS News*. Retrieved October 28, 2024, from https://www.cbsnews.com/news/taxes-what-happens-if-you-dont-file-tax-return/

Porto, N., & Collins, J. M. (2017). The role of refund expectations in savings: Evidence from volunteer income tax preparation programs in the United States. *Journal of Consumer Affairs*, *51*(1), 183–199. https://doi.org/10.1111/joca.12118

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rampell, C. (2024, June 14). How much did Congress lose by defunding the IRS? Way more than we thought. *The Washington Post*. Retrieved October 28, 2024, from https://www.washingtonpost.com/opinions/interactive/2023/irs-enforcement-costs-congress-funding/

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Schaupp, L. C., & Carter, L. (2010). The impact of trust, risk and optimism bias on e-file adoption. *Information Systems Frontiers*, *12*(3), 299–309. https://doi.org/10.1007/s10796-008-9138-8

Sela, A. (2018). Can computers be fair? How automated and human-powered online dispute resolution affect procedural justice in mediation and arbitration. *Ohio State Journal on Dispute Resolution*, *33*(1), 91–148. https://ssrn.com/abstract=3074311

Shestowsky, D. (2014). The psychology of procedural preference: How litigants evaluate legal procedures ex ante. *Iowa Law Review*, *99*(2), 637–710. https://ssrn.com/abstract=2378622

Simonsen, A., Scheel-Krüger, J., Jensen, M., Roepstorff, A., Møller, A., Frith, C. D., & Campbell-Meiklejohn, D. (2014). Serotoninergic effects on judgments and social learning of trustworthiness. *Psychopharmacology*, *231*(14), 2759–2769. https://doi.org/10.1007/s00213-014-3444-2

Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, *25*(3), 274–287. https://doi.org/10.1057/ejis.2015.17

Sousa, S., & Kalju, T. (2022). Modeling trust in COVID-19 contact-tracing apps using the Human–Computer Trust Scale: Online survey study. *JMIR Human Factors*, *9*(2), Article e33951. https://doi.org/10.2196/33951

Starkman, J. (2024, February 5). Beware of e-filing your tax return. *Wall Street Journal*. Retrieved October 28, 2024, from https://www.wsj.com/articles/beware-of-e-filing-your-tax-return-legal-trouble-for-error-privacy-risk-cyberattack-96d31111

Tizpaz-Niari, S., Wagner, M. R., Darian, S., Reed, K., & Trivedi, A. (2023). Metamorphic testing and debugging of tax preparation software. *Proceedings of the 45th International Conference on Software Engineering: Software Engineering in Society*. https://doi.org/10.48550/arXiv.2205.04998

Tyler, T. R. (1984). The role of perceived injustice in defendant's evaluations of their courtroom experience. *Law & Society Review*, *18*(1), 51–74. https://doi.org/10.2307/3053480

Tyler, T. R. (1990). *Why people obey the law*. Princeton University Press. https://press.princeton.edu/books/paperback/9780691126739/why-people-obey-the-law?srsltid=AfmBOooT2nzaNEll-m9ssQ_xsp83IE8TqEimeXP--jrziooAwIhcbuqt

Tyler, T. R. (1994). Governing amid diversity: The effect of fair decisionmaking procedures on the legitimacy of government. *Law & Society Review*, *28*(4), 809–831. https://doi.org/10.2307/3053998

Tyler, T. R. (2003). Procedural justice, legitimacy, and the effective rule of law. *Crime and Justice: A Review of Research*, *30*, 283–358. https://doi.org/10.1086/652233

Tyler, T. R. (2011). Trust and legitimacy: Policing in the USA and Europe. *European Journal of Criminology*, *8*(4), 254–266. https://doi.org/10.1177/1477370811411462

van Dijk, W. W., Goslinga, S., Terwel, B. W., & van Dijk, E. (2020). How choice architecture can promote and undermine tax compliance: Testing the effects of prepopulated tax returns and accuracy confirmation. *Journal of Behavioral and Experimental Economics*, *87*, Article 101574. https://doi.org/10.1016/j.socec.2020.101574

van Dijke, M., Gobena, L. B., & Verboon, P. (2019). Make me want to pay. A three-way interaction between procedural justice, distributive justice, and power on voluntary tax compliance. *Frontiers in Psychology*, *10*, Article 1632. https://doi.org/10.3389/fpsyg.2019.01632

Wei, Z., Zhao, Z., & Zheng, Y. (2019). Following the majority: Social influence in trusting behavior. *Frontiers in Neuroscience*, *13*, Article 89. https://doi.org/10.3389/fnins.2019.00089

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C. W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han, Y., Fu, C., Yin, Z., Liu, M., … Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *Innovation*, *2*(4), Article 100179. https://doi.org/10.1016/j.xinn.2021.100179

Received August 14, 2024
Revision received November 27, 2024
Accepted December 3, 2024 ■