nature research

https://doi.org/10.1038/s41562-020-00943-3

Supplementary information

The automatic influence of advocacy on lawyers and novices

In the format provided by the authors and unedited

The Automatic Influence of Advocacy on Lawyers and Novices

David E. Melnikoff & Nina Strohminger

Supplementary Materials

Supplementary Methods

Experiment 1

Introduction. The survey began with the following introduction:

> We are interested in the extent to which people can ignore goal-relevant visual imagery. To test this, you will complete a series of visual perception tasks in which you must ignore the image of a man who has been accused of financial fraud and embezzlement. You will also play a game called "Attorney at Law," which is based on the accused man's trial.
>
> Continue to learn about the accused man.

Description of Harrison. The story about Harrison began as follows:

> This is Harrison, a 52-year-old father of two. His friends and family describe him as a loving father and a dedicated husband. Harrison is the Chief Financial Officer of Skyworks Solutions Inc., a manufacturing company in the United States, and once a month he volunteers at his local soup kitchen to feed the homeless. Now, Harrison is on trial for financial fraud and embezzlement.

What happened next in the story constituted our manipulation of Harrison's guilt. Participants who learned that Harrison was innocent read the following:

> On October 15th, 2017, Harrison made a charitable donation of $1.5 million on behalf of Skyworks Solutions. Specifically, he donated $1.5 million to St. Jude's Children's Research Hospital to fund cancer research. Soon after Harrison made the charitable donation, Skyworks Solutions was hit by a cyber attack. The hackers manipulated the company's internal records, including the record of Harrison's charitable donation. The hackers altered the record to appear as though Harrison transferred the $1.5 million not to St. Jude's Children's Research Hospital, but to a personal offshore account. As a result, Federal investigators have charged Harrison with financial fraud and embezzlement, and Harrison faces up to 15 years in prison.

Participants who learned that Harrison was guilty received a different message:

> On October 15th, 2017, Harrison stole $1.5 million from his company, Skyworks Solutions. Specifically, he transferred $1.5 million to a personal offshore bank account. In an attempt to cover up his crime, Harrison altered the Skyworks Solutions' internal records to appear as though he transferred the $1.5 million not to a personal offshore account, but to St. Jude's Children's Research Hospital as a charitable donation on behalf of the company. However, other personnel at Skywork Solutions quickly discovered that the records had been altered. As a result, Federal investigators have charged Harrison with financial fraud and embezzlement, and Harrison faces up to 15 years in prison.

Description of advocacy roles. All participants received the following information immediately before they were randomly assigned to the role of defense attorney or prosecuting attorney:

> In "Attorney at Law," you will be either the prosecuting attorney or the defense attorney in Harrison's trial.
>
> In the United States justice system, fairness and justice are guaranteed by giving everyone due process. This means that everyone — guilty or innocent — has a trial with a prosecuting attorney and a defense attorney. It is the prosecuting attorney's sworn duty to disprove evidence provided by the defense, and it is the defense attorney's sworn duty to disprove evidence provided by the prosecution. The jury then considers all the evidence and decides whether or not to convict.

Participants assigned to the role of defense attorney then received the following message:

> You will be defending Harrison. Remember that all trials must be fair and balanced; even though Harrison is [innocent/guilty], all trials must follow due process. The jury will hear the negative, incriminating evidence from the prosecution, and counter-evidence from the defense (you). The jury will consider all of the evidence to decide whether or not to convict Harrison.
>
> Your job is to uphold the constitution of the United States by defending Harrison to the very best of your ability. You will do this by disproving as much of the prosecution's negative, incriminating evidence about Harrison as possible. Whether or not to convict Harrison is ultimately the decision of the jury. Your only job is to defend Harrison by disproving as much of the incriminating evidence as possible.
>
> We will give a $10 bonus to the 5 best defense attorneys. The 5 defense attorneys who disprove the most pieces of incriminating evidence will each receive a $10 bonus!

Participants assigned to the role of prosecuting attorney received a different message:

> You will be prosecuting Harrison. Remember that all trials must be fair and balanced; even though Harrison is innocent, all trials must follow due process. The jury will hear the positive, exonerating evidence from the defense, and counter-evidence from the prosecution (you). The jury will consider all of the evidence to decide whether or not to convict Harrison.
>
> Your job is to uphold the constitution of the United States by prosecuting Harrison to the very best of your ability. You will do this by disproving as much of the defense's positive, exonerating evidence about Harrison as possible. Whether or not to convict Harrison is ultimately the decision of the jury. Your only job is to prosecute Harrison by disproving as much of the exonerating evidence as possible.
>
> We will give a $10 bonus to the 5 best prosecuting attorneys. The 5 prosecuting attorneys who disprove the most pieces of exonerating evidence will each receive a $10 bonus!

Rules of Attorney at Law. Participants assigned to the role of defense attorney received the following rules for Attorney at Law:

> You will defend Harrison by disproving as many pieces of negative, incriminating evidence as possible.
>
> During the trial, pictures of people (including Harrison) will appear on your screen. When you see Harrison, the word "GUILTY" will appear under his photo. To disprove a piece of negative, incriminating evidence, you must eliminate the word "GUILTY" from your screen as fast as possible. To eliminate the word "GUILTY," you must press the SPACE BAR.
>
> To disprove a piece of negative, incriminating evidence, you must eliminate the word "GUILTY" within 2 seconds of seeing Harrison. If you take more than 2 seconds to eliminate the word "GUILTY," the trial will proceed and you will have lost the chance to disprove a piece of incriminating evidence.

Participants assigned to the role of prosecuting attorney received equivalent rules:

> You will prosecute Harrison by disproving as many pieces of positive, exonerating evidence as possible.
>
> During the trial, pictures of people (including Harrison) appear on the prosecuting attorney's screen. When the prosecuting attorney sees Harrison, the word "INNOCENT" appears under his photo. To disprove a piece of positive, exonerating evidence, the prosecuting attorney must eliminate the word "INNOCENT" from their screen as fast as possible. To eliminate the word "INNOCENT," the prosecuting attorney must press the SPACE BAR.
>
> To disprove a piece of positive, exonerating evidence, prosecuting attorneys must eliminate the word "INNOCENT" within 2 seconds of seeing Harrison. If the prosecuting attorney takes more than 2 seconds to eliminate the word "INNOCENT," the trial proceeds and the prosecuting attorney loses their chance to disprove a piece of exonerating evidence.

Vignettes for true self items. The five vignettes that described moral actions were (*i*) Harrison once helped an old woman carry her groceries up to her third-floor apartment; (*ii*) Harrison once bought a train ticket for someone who was stranded far from home with no money; (*iii*) Harrison once saved a young boy from downing by helping him up from the bottom of the pool; (*iv*) Harrison once received too much change from a cashier and immediately returned all of the money; (*v*) Harrison once spent an entire weekend volunteering for a soup kitchen. The five vignettes that described immoral actions were (*i*) Harrison once stole a stranger's credit card and used it to buy himself a new TV; (*ii*) Harrison once stole a bike that had been left unlocked; (*iii*) Harrison once cheated during a game of chess by moving one of his pieces while his opponent wasn't looking; (*iv*) Harrison once threw his dinner plate at his girlfriend for overcooking a meal; (*v*) Harrison once cut to the front of a long line at the post office.

Attention checks. Immediately after introducing Harrison, we asked two questions. First, we asked participants to indicate which of the following statements was true: "Harrison committed financial fraud and embezzlement" or "Harrison has been falsely accused of financial fraud and embezzlement." Second, we asked participants to indicate which of the following statements was

true: “Harrison made a charitable donation of $1.5 million to St. Jude’s Children’s Research Hospital” or “Harrison lied to make it appear as though he made a charitable donation of $1.5 million to St. Jude’s Children’s Research Hospital.” Immediately after assigning participants to a role in Attorney at Law, but before providing the rules, we asked, “Will you prosecute or defend Harrison?”. There were two response options: “Prosecute” and “Defend”. Immediately after presenting the rules for Attorney at Law, we asked, “During the trial, what will you do each time you see Harrison?”. There were two response options: “Eliminate the word “GUILTY” as fast as possible” or “Eliminate the word “INNOCENT” as fast as possible”.

Experiment 2

Descriptions of advocacy roles. All participants received the following information immediately before they were randomly assigned to the role of defense attorney or interrogator:

> In "Law and Order," you will have one of two roles: (1) Harrison's attorney or (2) an interrogator whose job is to make Harrison confess.
>
> In the United States justice system, fairness and justice are guaranteed by giving everyone due process. This means that everyone — guilty or innocent — has an attorney and is questioned by an interrogator. Regardless of whether the suspect is guilty or innocent, it is the attorney's sworn duty to prevent the suspect from confessing, and it is the interrogator's sworn duty to make the suspect confess.

Participants assigned to the role of defense attorney then received the following message:

> You have been assigned to the role of Attorney. Thus, your job is to keep Harrison out of jail by preventing him from confessing. There are three things you must do in order to prevent Harrison from confessing:
>
> 1. Make Harrison fear you. If Harrison fears you, he will do what you say, such as refusing to speak to the interrogator.
> 2. Make Harrison distrust the interrogator. The more Harrison distrusts the interrogator, the less willing he will be to share his secrets.
> 3. Make Harrison feel ashamed of his crime. The more ashamed Harrison feels of his crime, the less likely he will be to admit that he did it.
>
> We will award a $10 bonus to the 5 attorneys who make Harrison refuse to confess the fastest! The 5 interrogators who make Harrison refuse to confess the fastest will each receive a $10 bonus.

Participants assigned to the role of interrogator received a different message:

> You have been assigned to the role of Interrogator. Thus, your job is to get Harrison put in jail by manipulating him into confessing. There are three things you must do in order to get a confession:
>
> 1. Make Harrison like you. If Harrison likes you, he will want to be friends with you. Then you can offer him your friendship in exchange for his confession.
> 2. Make Harrison trust you. The more Harrison trusts you, the more willing he will be to tell you his secrets.

3. Make Harrison feel unashamed of his crime. The less ashamed Harrison feels, the likelier he will be to confess to his crime.

> We will award a $10 bonus to the 5 interrogators who make Harrison confess the fastest! The 5 interrogators with fastest confession times will each receive a $10 bonus.

Rules of Law and Order. Interrogators received the following instructions for Law and Order:

> Throughout the interrogation, you will be given opportunities to talk to Harrison. At each opportunity, you will choose 1 of 3 possible messages to communicate to Harrison. One message will make Harrison more likely to confess, one will make Harrison less likely to confess, and one will have no effect. Your goal is to make Harrison confess as fast as possible by choosing the right things to say to Harrison.

Defense attorneys received equivalent instructions:

> Throughout the interrogation, you will be given opportunities to talk to Harrison. At each opportunity, you will choose 1 of 3 possible messages to communicate to Harrison. One message will make Harrison more likely to confess, one will make Harrison less likely to confess, and one will have no effect. Your goal is to make Harrison refuse to confess as fast as possible by choosing the right things to say to Harrison.

Attention checks. We asked the same two questions about Harrison's guilt used in Experiment 1. Immediately after assigning participants to the role of defense attorney or interrogator, but before presenting the rules for Law and Order, we asked two questions to ensure that participants understood their role. First, we asked participants to select three phrases from a list of six that would be good for them to say during the interrogation. The six phrases were (*i*) "Everyone makes mistakes. That doesn't make you a bad person." (*ii*) "I'm on your side, Harrison. You can trust me." (*iii*) "It says here that you're the CFO of Skyworks. Very impressive!" (*iv*) "If you admit what you did, your reputation will be ruined forever." (*v*) "Right now, assume that everyone is out to get you. Trust no one." (*vi*) "Don't talk unless I tell you to. You may be a CFO, but here, I'm in charge." Phrases *i-iii* were good for interrogators to say, and phrases *iv-vi* were good for defense attorneys to say. The phrases were presented in randomized order. After participants selected three phrases, we asked them to indicate which of the following statements was true: "I will try to get Harrison put in jail by manipulating him into confessing" or "I will try to keep Harrison out of jail by preventing him from confessing".

Immediately after presenting the instructions for Law and Order, we asked, "During the interrogation, what will you do each time you have an opportunity to talk to Harrison?" There were three response options: (*i*) "Choose the message that will make Harrison more likely to confess" (*ii*) "Choose the message that will make Harrison less likely to confess" or (*iii*) "Choose the message that will have no effect".

Experiment 3

Introduction. Participants were introduced to Experiment 3 as follows:

> We are interested in people's ability to form accurate, unbiased opinions. To help us explore this topic, you will play a game called "Attorney at Law," which is based on the trial of a man

who was accused of a crime. In order to win the game, you must evaluate the accused man as accurately as possible.

Continue to learn about the accused man.

Attention checks. The attention checks in Experiment 3 were identical to those from Experiment 1 with two exceptions. First, we eliminated the question asking what participants should do during the trial, since specific rules for Attorney at Law were omitted. Second, we added an attention check immediately after the message informing participants that their role as prosecuting attorney or defense attorney might bias their evaluations, and that avoiding such bias is imperative. Specifically, we asked participants to indicate which of the following statements was true: "In order to succeed, I must form fair and unbiased opinions of Harrison" or "In order to succeed, I must form unfair and biased opinions of Harrison".

Experiment 4

Introduction. We introduced the study in one of two ways depending on whether participants were to be a prosecuting attorney or learner. The introduction for prosecuting attorney was as follows:

> We are studying how task-irrelevant information affects memory. To help us, you will (*i*) learn the rules of a game called Attorney at Law, and then (*ii*) learn a series of irrelevant facts. We are interested in whether you remember the rules of Attorney at Law after you have learned the series of irrelevant facts.
>
> To demonstrate whether you remember the rules of Attorney at Law, you will actually play Attorney at Law at the end of the survey. This is because we are interested in people's ability to remember rules that they plan to follow.

The introduction for learners was as follows:

> We are studying how task-irrelevant information affects memory. To help us, you will (*i*) learn the rules of a game called Attorney at Law, and then (*ii*) learn a series of irrelevant facts. We are interested in whether you remember the rules of Attorney at Law after you have learned the series of irrelevant facts.
>
> You will NOT play Attorney at Law at any point. Instead, you will simply recall the rules of Attorney at Law at the end of the survey. This is because we are interested in people's ability to remember rules that they do not plan to follow at any point in the future.

Vignettes for true self items. The three vignettes that described moral actions were (*i*) Harrison once helped an old woman carry her groceries up to her third-floor apartment; (*ii*) Harrison once received too much change from a cashier and immediately returned all of the money; (*iii*) Harrison once spent an entire weekend volunteering for a soup kitchen. The three vignettes that described immoral actions were (*i*) Harrison once stole a stranger's credit card and used it to buy himself a new TV; (*ii*) Harrison once stole a bike that had been left unlocked; (*iii*) Harrison once cheated during a game of chess by moving one of his pieces while his opponent wasn't looking.

Attention checks. To ensure that participants knew whether they would or would not actually play Attorney at Law, we asked the following question immediately after introducing the study: "When will you play Attorney at Law?" The response options were: (*i*) At the beginning of the survey; (*ii*) At the end of the survey; (*iii*) Never. To ensure that participants knew that Harrison was innocent, we asked the same two questions about Harrison's guilt used in Experiment 1.

Immediately after randomly assigning participants to the role of prosecuting attorney or learner, but before presenting the rules for Attorney at Law, we asked one of two questions depending on the role to which participants were assigned. We asked prosecuting attorneys, "When you play Attorney at Law, you will have one of two roles. What is your role?" and we asked learners, "Instead of playing Attorney at Law, you are learning the rules for one of two roles. Which role are you learning about?" Both questions included two response options: Prosecuting Attorney or Defense Attorney.

Immediately after providing the rules for Attorney at Law, we asked one of two question depending on the condition to which participants were assigned. We asked prosecuting attorneys, "During the trial, what will you do each time you see Harrison?" and we asked learners, "During the trial, what must prosecuting attorneys do each time they see Harrison?" Both questions had two response options: "Eliminate the word "GUILTY" as fast as possible" or "Eliminate the word "INNOCENT" as fast as possible".

Experiment 6

Procedure. Participants began by learning that we are studying how power influences people's ability to compete or cooperate with others. To explore this question, the cover story explained, participants would complete two different tasks. In the first task they would have power over someone, and in the second they would cooperate or compete with that person in a game. Participants responded to a multiple-choice item asking them to describe the two types of tasks they would be completing (power and cooperation vs. power and competition) — the first attention check.

Next, participants learned that they would play a game called Politico with another participant, whose username was dm88. Participants were assigned to one of two roles in Politico: campaign manager or candidate. Campaign managers learned that Politico involves two roles: (*i*) a political candidate running for office, who can earn $5.00 by getting more votes than their opponent (another participant), and (*ii*) that candidate's campaign manager. These participants learned that they would be the campaign manager and that dm88 would be the candidate. Campaign managers then learned that they must advocate for people to vote for dm88 over dm88's opponent. As the incentive, campaign managers learned that they would receive $5.00 if dm88 were to get more votes than their opponent, but would receive no bonus if dm88 were to get fewer votes.

Participants assigned to the role of candidate were informed that they were a politician running against dm88, and must advocate for people to *not* vote for dm88. As incentive, candidates learned that they would earn $5.00 if they were to receive more votes than dm88, but if they receive fewer votes, they would earn $0.00 (and dm88 would earn the $5.00). Participants responded to three multiple choice items asking (*i*) about their role, (*ii*) what will happen if they/dm88 get more votes than their opponent, and (*iii*) what will happen if they/dm88 get fewer votes than their opponent — the second, third, and fourth attention checks.

Next, we told campaign managers and candidates how Politico is played to ensure that they had similar expectations of what their advocacy would entail. Campaign managers learned that they would create a political ad advocating for people to vote for dm88. Campaign managers would write

a slogan (e.g., "Vote dm88 for a Better Future") to be overlaid on one of 5 images of their choice. They saw the 5 available images, all of which were stereotypical of campaign ads (the American flag, lady justice, people holding hands, someone summiting a mountain, and pairs of hands making a giving gesture). Campaign managers learned that their ad would be shown to a naïve group of participants alongside an ad for dm88's opponent (designed by the opponent's campaign manager), and that those naïve participants would vote for dm88 or dm88's opponent on the basis of the ads.

Participants assigned to the role of candidate also learned that they must make a political ad. Theirs, however, would advocate for people to vote *against* dm88. Thus, candidates would write a slogan (e.g., "Don't vote for dm88") to be overlaid on one of five images stereotypical of political attack ads (a crying baby, a homeless man, a burning American flag, a businessman with the head of a pig, or a man putting on a Pinocchio mask). These participants learned that their ad would be shown to a naïve group of participants alongside an ad that dm88 had created, and that those naïve participants would vote for them of dm88 on the basis of the ads.

By this point, participants (*i*) were incentivized to advocate for people to vote for or against dm88, and (*ii*) had similar expectations for how their advocacy would proceed. Next, participants proceeded to what we called the "power task". In line with the cover story, participants learned that the purpose of the power task is to give them power over dm88, allowing us to see how power influences cooperativeness (if participants were in the campaign manager condition) or competitiveness (if participants were in the candidate condition). The power task, we explained, is completely unrelated to Politico — its only purpose is to give participants the experience of having power over dm88.

In the power task, participants learned that dm88 had played a game where it was possible to cheat for monetary gain. In this game, players earn money for each trivia question they answer correctly. However, they are asked to score themselves, giving players the opportunity to cheat by reporting incorrect answers as correct. Participants saw dm88's answers alongside the correct answers, and whether dm88 reported each answer as correct or incorrect. We explained to participants that they would have the power to reward dm88 for honestly reporting incorrect answers, and to punish dm88 for dishonestly reporting correct answers. We presented two multiple choice items asking participants to identify (*i*) the circumstances under which they could reward dm88, and (*ii*) the circumstances under which they could punish dm88 — the fifth and sixth attention checks. As the seventh and final attention check, participants were asked, for each of dm88's answers, whether dm88 was (*i*) incorrect and lied about it, (*ii*) incorrect and honest about it, or (*iii*) correct.

Dm88 answered seven questions, three correctly and four incorrectly. Dm88 dishonestly reported half of the incorrect answers. For the incorrect answers, participants were asked the same true self questions as in Experiments 1–4: "In your opinion, what aspect of dm88's personality caused dm88 to be [honest/dishonest] about their answer?" and "When dm88 was [honest/dishonest] about their answer, to what extent was dm88 being true to the deepest, most essential aspects of dm88's being?" To explore whether advocacy incentives alter overt behavior, we also had participants choose an amount of money— from zero cents to five cents—to add to dm88's earnings (for each honest answer), or to deduct from dm88's earnings (for each dishonest answer). If mere incentives to advocate alter true self beliefs and overt behavior, then candidates, relative to campaign managers, should have more negative beliefs about dm88's true self, and should punish dm88's dishonesty more than they reward dm88's honesty.

After measuring true self beliefs and overt behavior, participants created their political ad by writing a slogan and selecting the image with which they wanted to pair it. Participants then completed a demographics survey and were debriefed. Bonus payments were administered by having a separate group of participants vote between randomly generated pairs of ads. For each pair, the

authors of the ad that received the most votes earned a $5.00 bonus. Because participants did not actually advocate until all dependent measures had been collected, systematic differences between candidates and campaign managers can only be attributed to incentives.

Experiment 7

Description of Harrison. All participants received the following description of Harrison:

> This is Harrison, a 52-year-old father of two. His friends and family describe him as a loving father and a dedicated husband. Harrison is the Chief Financial Officer of Skyworks Solutions Inc., a manufacturing company in the United States, and once a month he volunteers he volunteers at a local soup kitchen.
>
> Now, Harrison is on trial after being accused of financial fraud and embezzlement. Specifically, Harrison has been accused of embezzling $1.5 million. If found guilty, Harrison faces up to 19 years in prison.

Attention checks. Experiment 7 included all of the attention checks from Experiment 1 with the following exceptions. First, we eliminated the two items from Experiment 1 that asked whether Harrison was innocent or guilty, and replaced them with a single item that asked participants to indicate which of the following statements was true: (*i*) Harrison is definitely innocent of theft, (*ii*) Harrison is definitely guilty of theft, or (*iii*) It is unknown whether Harrison is innocent or guilty of theft. Second, we included an additional attention check, presented immediately after informing participants that they must correctly guess whether Harrison is innocent or guilty. Specifically, we asked participants to indicate which of the following statements was true: (*i*) "In order to [defend/prosecute] Harrison, I must correctly guess whether he is innocent or guilty" or (*ii*) "I will [defend/prosecute] Harrison whether or not I correctly guess whether he is innocent or guilty."

Experiment 8

Description of Harrison. All participants received the following description of Harrison:

> This is Harrison, a 52-year-old American male. Harrison was accused of robbing a jewelry store, and is awaiting trial.
>
> A security camera at the store shows a man stealing over $30,000 worth of jewelry. However, the footage is blurry, so the man cannot be identified with certainty. Nonetheless, when investigators compared the video footage with images of Harrison, they were confident enough that they had a match to issue an arrest warrant.
>
> After Harrison's arrest, the video footage was filtered through image de-noising algorithms (to make it less blurry) and face recognition algorithms (to compute the exactly probability that the person in the image is Harrison). These algorithms determined that there is a 60% chance that Harrison is the man in the video.
>
> If found guilty, Harrison faces up to 19 years in prison.

Attention checks. The attention checks in Experiment 8 were identical those from Experiment 7, with one addition: Immediately after describing Harrison, we asked participants, "What is the probability that the man in the security footage is Harrison?" Participants answered this question by typing a percentage from 0% to 100%.

Experiment 9

Description of advocacy role. Participants randomly assigned to advocate for Harrison advocates read the following:

> For the purposes of this survey, imagine that you have been retained by Harrison West, a Texas man, who is being sued by Ineos USA, a chemical company, for publishing the company's trade secrets. Additional details are provided below:

Participants randomly assigned to advocate for Ineos advocates read:

> For the purposes of this survey, imagine that you have been retained by Ineos USA, a chemical company, to bring suit against Harrison West, a Texas man, for publishing the company's trade secrets. Additional details are provided below:

All participants then received the following information about the case:

> Ineos USA, a chemical company, is seeking $8.5 million in damages and an injunction against Harrison West, a Texas man, for publishing the company's trade secrets. West is a former employee who approached a current Ineos employee, asking him to edit a book that West had written and subsequently published. The current employee reported this to his boss. Ineos's in-house counsel subsequently filed a lawsuit, alleging that the book contains proprietary information in violation of the state's Information Theft Liability Act, as well as a breach of West's duty of loyalty.

Incentivization. As incentive, we prompted participants to see the process of advocating for Ineos or Harrison as personally fulfilling. We did this by having participants write free responses to the following three questions: (*i*) "What aspect of being [Ineos'/West's] lawyer would give you the most personal fulfillment?" (*ii*) "What is one of the most important tasks that you would accomplish during your first month as [Ineos'/West's] lawyer?" and (*iii*) "Please describe how the task that you described above would help you achieve your goal of [litigating against/defending] West." Participants then reported their beliefs about Harrison's true self using the same measure from Experiment 1. Next, participants were asked about Harrison's factual guilt ("Do you think that West in fact violated the Theft Liability Act?"), liability ("Do you think that West should be liable for damages to Ineos USA?"), and support for an injunction ("Do you think that an injunction should be enforced against West?"). Participants responded to all three items using a 9-point scale from 0 (Definitely Yes) to 8 (Definitely No). Participants also answered the following question about damages: "Ineos is seeking $8.5 million in damages from West. If West was ordered to pay damages, what percentage of the $8.5 million do you think he should have to pay?" Participants answered this question by providing a percentage from 0% from 100%. Finally, participants completed a demographics survey and were debriefed without ever advocating for either side.

Experiment 10

Introduction. Participants who were to imagine being retained by Harrison West were introduced to the experiment as follows:

> We are interested in the extent to which legal training improves people's ability to form accurate beliefs about the guilt or innocence of people accused of wrongdoing.
>
> To help us, you will imagine that you have been retained by Harrison West, a Texas man, who is being sued by Ineos USA, a chemical company, for allegedly stealing trade secrets (names and identifying details have been changed to protect the privacy of the individuals involved in the case). You will receive information about the case, and West himself. Then you will guess whether West did, in fact, steal trade secrets from Ineos USA, which would make West liable for any resulting damages.
>
> The facts of the case are known, allowing us to assess the accuracy of your answer. If you correctly guess whether West stole trade secrets from Ineos USA, you will be entered into a lottery for a chance to win $200 in bonus money.

Participants who were to imagine being retained by Ineos received a different introduction:

> We are interested in the extent to which legal training improves people's ability to form accurate beliefs about the guilt or innocence of people accused of wrongdoing.
>
> To help us, you will imagine that you have been retained by Ineos USA, a chemical company, who brought suit against Harrison West, a Texas man, for allegedly stealing trade secrets (names and identifying details have been changed to protect the privacy of the individuals involved in the case). You will receive information about the case, and West himself. Then you will guess whether West did, in fact, steal trade secrets from Ineos USA, which would make West liable for any resulting damages.
>
> The facts of the case are known, allowing us to assess the accuracy of your answer. If you correctly guess whether West stole trade secrets from Ineos USA, you will be entered into a lottery for a chance to win $200 in bonus money.

Case briefing. All participants received the following information about the case:

> Ineos USA, a chemical company, is seeking $8.5 million in damages and an injunction against Harrison West, a Texas man, for allegedly stealing the company's trade secrets. West is a former Ineos employee who discussed a book he wrote, and subsequently published, with a current Ineos employee. The current employee reported this information to an Ineos supervisor. Ineos's in-house counsel subsequently filed a lawsuit, alleging that the book contains proprietary information in violation of the state's Theft Liability Act, as well as a breach of West's duty of loyalty.
>
> Under Texas Penal Code § 31.05(a), a trade secret is "the whole or any part of any scientific or technical information, design, process, procedure, formula, or improvement that has value and that the owner has taken measures to prevent from becoming available to persons other

than those selected by the owner to have access for limited purposes." A person is guilty of stealing a trade secret when he "knowingly (1) steals a trade secret, (2) makes a copy of an article representing a trade secret, or (3) communicates or transmits a trade secret." Id. at § 31.05(b). In order for a defendant to be liable under § 31.05, he must have taken trade secrets "without the owner's effective consent." Id.. "Effective consent" includes consent by a person legally authorized to act for the owner.

Supplementary Results

Experiment 1

True self beliefs: Continuous. In this and every subsequent experiment, we used the lme4 package in R to run linear mixed effects models (LMMs) testing the effect of Role (prosecuting vs. defense attorney), Guilt (innocent vs. guilty) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs. Our initial model included fixed effects for all three factors and their interaction terms, plus random intercepts for participant and vignette. Adding a random slope for Action Type improved model fit, $\chi^2(2) = 2502.4$, $p < .001$, as did adding a second random slope for Role, $\chi^2(2) = 8.58$, $p = .014$. Adding a third random slope for Guilt had no effect, $\chi^2(3) = 3.6$, $p = .308$. Accordingly, we retained only the random slopes for Action Type and Role.

Next, we eliminated the fixed effect associated with the three-way interaction. This did not reduce the quality of the fit, $\chi^2(1) = .20$, $p = .652$, so we dropped the three-way interaction term from the model. Then we eliminated all fixed effects associated with the two-way interactions. This reduced the quality of the fit, $\chi^2(3) = 50.8$, $p < .001$, so we retained the two-way interaction terms.

The final model included fixed effects for all three factors and their two-way interactions, random intercepts for participant and vignette, and random slopes for Action Type and Role. The results of this model are presented in the main text.

In addition to LMMs, we analyzed the continuous measure of true self beliefs using mixed-effects analysis of variance (ANOVA). Specifically, we computed the mean score for moral and immoral actions separately, and treated the resulting scores as two levels of a within-subjects factor called Action Type. In this and every subsequent experiment, we report mixed-effects ANOVAs as a robustness check, and because they were included in our preregistered analysis plans for Experiments 3–7. Mixed-effects ANOVAs ignore vignette-specific variance, and therefore yield less reliable estimates than LMMs [1], which is why we report both types of analyses. But this distinction ends up not making any material difference to our conclusions, as mixed-effects ANOVAs and LMMs yielded similar results across all studies.

The mixed-effects ANOVA revealed a significant Guilt x Action Type interaction, $F(1, 382) = 16.45$, $p < .001$. Participants who learned that Harrison was innocent believed that Harrison's true self was significantly more moral ($M = 6.99$, 95% CI [6.73, 7.25]) than immoral ($M = 5.35$, 95% CI [5.02, 5.69]; $F(1, 382) = 43.26$, $p < .001$). In contrast, participants who learned that Harrison was guilty did not believe that Harrison's true self was more ($M = 6.04$, 95% CI [5.75, 6.34]) than immoral ($M = 5.85$, 95% CI [5.51, 6.2]; $F(1, 382) = .56$, $p = .455$).

We also found a significant Role x Action Type interaction, $F(1, 382) = 66.11$, $p < .001$. Defense attorneys believed that Harrison's true self was significantly more moral ($M = 7.17$, 95% CI [6.94, 7.41]) than immoral ($M = 4.79$, 95% CI [4.44, 5.13]; $F(1, 382) = 92$, $p < .001$). Conversely, prosecuting attorneys believed that Harrison's true self was significantly less moral ($M = 5.85$, 95% CI [5.55, 6.16]) than immoral ($M = 6.44$, 95% CI [6.15, 6.73]; $F(1, 382) = 5.36$, $p = .021$).

True self beliefs: Categorical. In this and every subsequent experiment, we used the lme4 package in R to run binomial generalized linear mixed effects models (GLMMs) to test the effects of Role, Action Type, and Guilt on the categorical measure of true self beliefs. 95% confidence intervals around group means were estimated from 1000 nonparametric bootstraps.

In Experiment 1, our initial model included fixed effects for all 3 factors and their interaction terms, plus random intercepts for participant and vignette. Adding a random slope for Action Type improved model fit, $\chi^2(2) = 1117.5$, $p < .001$, and no additional improvement resulted

from further adding a random slope for Role, $\chi^2(2) = 1.25$, $p = .536$, or Guilt, $\chi^2(2) = 1.22$, $p = .543$. Accordingly, the only random slope we retained was the one for Action Type. Next, we eliminated the fixed effect associated with the three-way interaction. This did not reduce the quality of the fit, $\chi^2(1) = .83$, $p = .363$, so we dropped the three-way interaction term from the model. Then we eliminated all fixed effects associated with the two-way interactions. This reduced the quality of the fit, $\chi^2(3) = .83$, $p < .001$, so we retained the two-way interaction terms.

The final model included fixed effects for all three factors and their two-way interaction terms, random intercepts for participant and vignette, and a random slope for Action Type. This model revealed a significant Guilt x Action Type interaction ($b = 2.89$, $SE = .64$, $z = 3.58$, $p < .001$). Participants who learned that Harrison was innocent were significantly more likely to associate Harrison's true self with moral actions ($M = 69.42\%$, 95% CI [68.44%, 70.44%]) than with immoral actions ($M = 46.54\%$, 95% CI [45.34%, 47.85%]), $b = 1.95$, $SE = .57$, $z = 3.44$, $p < .001$. In contrast, participants who learned that Harrison was guilty were not more likely to associate Harrison's true self with moral actions ($M = 54.29\%$, 95% CI [53.23%, 55.45%]) than with immoral actions ($M = 57.7\%$, 95% CI [56.52%, 58.92%]; $b = .34$, $SE = .57$, $z = .58$, $p = .559$).

This model also revealed a significant Role x Action Type interaction ($b = 5.57$, $SE = .68$, $z = 8.22$, $p < .001$). Defense attorneys were significantly more likely to associate Harrison's true self with moral actions ($M = 77.35\%$, 95% CI [76.54%, 78.12%]) than with immoral actions ($M = 33.26\%$, 95% CI [32.25%, 34.23%]; $b = 3.59$, $SE = .59$, $z = 6.11$, $p < .001$). Conversely, prosecuting attorneys were significantly less likely to associate Harrison's true self with moral actions ($M = 47.7\%$, 95% CI [45.63%, 47.8%]) than with immoral actions ($M = 69.37\%$, 95% CI [68.36%, 70.38%]; $b = 1.98$, $SE = .58$, $z = 3.43$, $p < .001$).

In addition to GLMMs, we analyzed the categorical measure of true self beliefs using mixed-effects ANOVAs. We separately computed the total number of moral and immoral actions attributed to Harrison's true self, and treated the resulting scores as two levels of a within-subjects factor called Action Type. As with the continuous measure of true self beliefs, we analyzed the categorical measure with ANOVA as a robustness check, and because ANOVAs were included in our preregistered analysis plans for Experiments 3–7. In every study, ANOVAs and GLMMs yielded similar results.

The mixed-effects ANOVA revealed a significant Guilt x Action Type interaction, $F(1, 382) = 20.36$, $p < .001$. Participants who learned that Harrison was innocent associated Harrison's true self with significantly more moral actions ($M = 3.43$, 95% CI [3.19, 3.67]) than immoral actions ($M = 2.32$, 95% CI [2.04, 2.6]), $F(1, 382) = 26$, $p < .001$). In contrast, participants who learned that Harrison was guilty did not associate Harrison's true self with more moral actions ($M = 2.64$, 95% CI [2.36, 2.93]) than immoral actions ($M = 2.94$, 95% CI [2.64, 3.24]; $F(1, 382) = 1.77$, $p = .184$).

We also found a significant Role x Action Type interaction, $F(1, 382) = 75.92$, $p < .001$. Defense attorneys associated Harrison's true self with significantly more moral actions ($M = 3.71$, 95% CI [3.47, 3.94]) than immoral actions ($M = 1.91$, 95% CI [1.63, 2.19]; $F(1, 382) = 67.65$, $p < .001$). Conversely, prosecuting attorneys associated Harrison's true self with significantly fewer moral actions ($M = 2.36$, 95% CI [2.09, 2.63]) than immoral actions ($M = 3.37$, 95% CI [3.1, 3.63]); $F(1, 382) = 20.31$, $p < .001$).

Surface self beliefs: Categorical. In this and every subsequent study, we ran binomial GLMMs to test the effect of Role, Action Type, and Guilt on the likelihood of attributing actions to Harrison's surface self. We did so to ensure that what appear to be effects of advocacy incentives on true self beliefs are not results of a more general response bias (i.e. a general tendency to attribute more or less actions to the actor). This alternative explanation was ruled out in every experiment; the

relationship between advocacy incentives and surface self beliefs was always categorically different than the relationship between advocacy incentives and true self beliefs.

In Experiment 1, our initial model included fixed effects for all three factors and their interaction terms, plus random intercepts for participant and vignette. Adding a random slope for Action Type improved model fit, $\chi^2(2) = 925.07$, $p < .001$. This model was not improved further by adding a random slope for Role, $\chi^2(2) = .703$, $p = .704$, or Guilt, $\chi^2(2) = .353$, $p = .838$. Accordingly, the only random slope we retained was for Action Type. Next, we eliminated the fixed effect associated with the three-way interaction. This reduced the quality of the fit, $\chi^2(1) = 4.84$, $p = .028$, so we kept the three-way interaction term in the model.

The final model included fixed effects for all three factors and their interaction terms, random intercepts for participant and vignette, and a random slope for Action Type. This model revealed a significant Role x Action Type x Guilt interaction ($b = 2.54$, $SE = 1.02$, $z = 2.49$, $p = .013$). Specifically, the Role x Action Type interaction was significant at both levels of Guilt, but stronger among participants who learned that Harrison is guilty ($b = 5.87$, $SE = .85$, $z = 6.87$, $p < .001$) versus those who learned that Harrison is innocent ($b = 3.36$, $SE = .80$, $z = 4.22$, $p < .001$).

Focusing first on participants who learned that Harrison is guilty, we examined the effect of Action Type separately for prosecuting attorneys and defense attorneys. We found that defense attorneys were significantly less likely to associate Harrison's surface self with moral actions (M = 23.7%, 95% CI [22.82%, 24.47%]) than with immoral actions (M = 48.83%, 95% CI [47.74%, 49.9%]; $b = 1.94$, $SE = .69$, $z = 2.8$, $p = .005$). Conversely, prosecuting attorneys were significantly more likely to associate Harrison's surface self with moral actions (M = 56.54%, 95% CI [55.53%, 57.57%]) than with immoral actions (M = 12.4%, 95% CI [11.77%, 13%]; $b = 1.94$, $SE = .69$, $z = 2.8$, $p = .005$). For participants who learned that Harrison is innocent, defense attorneys were significantly less likely to associate Harrison's surface self with moral actions (M = 15.96%, 95% CI [15.27%, 16.66%]) than with immoral actions (M = 44.42%, 95% CI [43.46%, 45.57%]; $b = 2.37$, $SE = .65$, $z = 3.65$, $p < .001$). In contrast, prosecuting attorneys were not less likely to associate Harrison's true self with moral actions (M = 38.52%, 95% CI [37.59%, 39.53%]) than with immoral actions (M = 27.61%, 95% CI [26.61%, 28.56%]; $b = .99$, $SE = .68$, $z = 1.45$, $p = .147$).

For the ANOVA, we separately computed the total number of moral and immoral actions attributed to Harrison's surface self, and treated the resulting scores as two levels of a within-subjects factor called Action Type. This test revealed a significant Guilt x Role x Action Type interaction, $F(1, 382) = 4.02$, $p = .046$. We decomposed this interaction by exploring the Role x Action Type interaction separately when Harrison was innocent versus guilty. When Harrison was innocent, the Role x Action Type interaction was significant ($F(1, 382) = 5.0$, $p = .026$) such that defense attorneys believed that Harrison's surface self was significantly less moral (M = .86, 95% CI [.62, 1.1]) than immoral (M = 2.1, 95% CI [1.76, 2.45]; $F(1, 382) = 20.22$, $p < .001$), whereas prosecuting attorneys did not believe that Harrison's surface self was less moral (M = 1.91, 95% CI [1.54, 2.28]) than immoral (M = 1.44, 95% CI [1.08, 1.8]; $F(1, 382) = 2.49$, $p = .116$). When Harrison was guilty, the Role x Action Type interaction was significant ($F(1, 382) = 6.38$, $p = .012$), such that defense attorneys believed that Harrison's surface self was significantly less moral (M = 1.32, 95% CI [.96, 1.68]) than immoral (M = 2.3, 95% CI [1.86, 2.74]; $F(1, 382) = 10.52$, $p = .001$), and prosecuting attorneys believed that Harrison's surface self was significantly more moral (M = 2.64, 95% CI [2.26, 3.02]) than immoral (M = .73, 95% CI [.48, .99]; $F(1, 382) = 43.65$, $p < .001$).

Experiment 2

True self beliefs: Continuous. To test the effect of Role and Action Type on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type, $\chi^2(2) = 2040.7$, $p < .001$, but was not further improved by adding a second random slope for Role, $\chi^2(2) = 1.01$, $p = .603$. Accordingly, the only random slope retained was the one for Action Type.

The final model included fixed effects for Role, Action Type, and their interaction term; random intercepts for participant and vignette; and a random slope for Action Type. The results of this model are presented in the main text.

The mixed-effects ANOVA revealed a significant Role x Action Type interaction ($F(1, 301) = 11.22$, $p < .001$), such that interrogators believed that Harrison's true self was significantly more immoral ($M = 7.13$, 95% CI [6.88, 7.38]) than moral ($M = 5.69$, 95% CI [5.39, 5.99]; $F(1, 301) = 35.84$, $p < .001$). In contrast, defense attorneys did not believe that Harrison's true self was more immoral ($M = 6.4$, 95% CI [6.07, 6.74]) than moral ($M = 6.15$, 95% CI [5.84, 6.47]; $F(1, 301) = .92$, $p = .338$).

True self beliefs: Categorical. To test the effect of Role and Action on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for Role, Action Type, and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type, $\chi^2(2) = 984.37$, $p < .001$, and no additional improvement resulted from further adding a second random slope for Role, $\chi^2(2) = .31$, $p = .858$. Accordingly, the only random slope we retained was the one for Action Type.

The final model included fixed effects for Role, Action Type, and their interaction; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction ($b = 1.99$, $SE = .81$, $z = 2.47$, $p = .014$). Interrogators were significantly more likely to associate Harrison's true self with immoral actions ($M = 81.01\%$, 95% CI [79.94%, 81.94%]) than moral actions ($M = 39.47\%$, 95% CI [38.27%, 40.65%]; $b = 4.2$, $SE = .74$, $z = 5.65$, $p < .001$). Defense attorneys had a weaker, but still significant tendency to associate Harrison's true self with immoral actions ($M = 70.16\%$, 95% CI [68.97%, 71.44%]) at a higher rate than moral actions ($M = 46.98\%$, 95% CI [45.84%, 48.13%]; $b = 2.2$, $SE = .75$, $z = 2.96$, $p = .003$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 301) = 5.98$, $p = .015$. Interrogators associated Harrison's true self with significantly more immoral actions ($M = 3.96$, 95% CI [3.71, 4.21]) than moral actions ($M = 2.13$, 95% CI [1.84, 2.41]), $F(1, 301) = 55.97$, $p < .001$. Defense attorneys had a weaker, but still significant pattern of associating Harrison's true self with more immoral actions ($M = 3.43$, 95% CI [3.11, 3.76]) than moral actions ($M = 2.48$, 95% CI [2.16, 2.8]; $F(1, 301) = 12.7$, $p < .001$).

Surface self beliefs: Categorical. To test effects of Role and Action Type on surface self beliefs, we started with a GLMM that included fixed effects for Role, Action Type, and their interaction term, plus random intercepts for participant and vignette. Model fit was improved by adding a random slope for Action Type, $\chi^2(2) = 80.19$, $p < .001$, and was further improved by adding a second random slope for Role, $\chi^2(2) = 11.9$, $p = .003$. Accordingly, we retained the slope for both Action Type and Role.

The final model included fixed effects for Action Type, Role, and their interaction term; random intercepts for participant and vignette; and random slopes for Action Type and Role. This model revealed a significant Role x Action Type interaction ($b = .76$, $SE = .34$, $z = 2.26$, $p = .024$). Interrogators were significantly more likely to associate Harrison's surface self with moral actions ($M = 48.15\%$, 95% CI [47.24%, 48.98%]) than immoral actions ($M = 22.19\%$, 95% CI [21.53%,

22.84%]; b = 1.48, SE = .67, z = 2.22, p = .027). In contrast, defense attorneys were not more likely to associate Harrison's surface self with moral actions (M = 42.38%, 95% CI [41.62%, 43.17%]) than immoral actions (M = 28.28%, 95% CI [27.67%, 28.88%]; b = .71, SE = .48, z = 1.49, p = .136).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 301)$ = 6.1, p = .014. Interrogators associated Harrison's surface self with significantly more moral actions (M = 2.75, 95% CI [2.47, 3.03]) than immoral actions (M = .75, 95% CI [.53, .97]; $F(1, 301)$ = 76.32, p < .001). Defense attorneys had a weaker, but still significant pattern of associating Harrison's surface self with more moral actions (M = 2.34, 95% CI [2.05, 2.67]) than immoral actions (M = 1.19, 95% CI [.90, 1.49]); $F(1, 301)$ = 21.97, p < .001).

Experiment 3

True self beliefs: Continuous. To test the effect of Role (prosecuting attorney vs. defense attorney), Guilt (guilty vs. innocent), and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for all three factors and their interaction terms, plus random intercepts for participant and vignette. Adding a random slope for Action Type improved model fit ($\chi^2(2)$ = 1484.8, p < .001). No further improvement came from adding a second random slope for Guilt ($\chi^2(2)$ = .86, p = .65) or Role ($\chi^2(1)$ = 3.39, p = .065). Accordingly, we retained only the random slope for Action Type.

Next, we eliminated the fixed effect associated with the three-way interaction. This did not reduce the quality of the fit ($\chi^2(1)$ = 1.85, p = .174), so we dropped the three-way interaction term from the model. Then we eliminated all fixed effects associated with the two-way interactions. This reduced the quality of the fit, $\chi^2(3)$ = 30.61, p < .001, so we retained the two-way interaction terms.

The final model included fixed effects for Role, Guilt, Action Type, and their two-way interaction terms; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a nonsignificant Guilt x Action Type interaction (b = .14, SE = .27, $t(344)$ = .52, p = .60). Participants believed that Harrison's true self more moral than immoral irrespective of whether he was described as innocent (M_{moral} = 6.78, 95% CI [6.43, 7.13]; $M_{immoral}$ = 5.95, 95% CI [5.57, 6.34], b = .83, SE = .27, $t(34.49)$ = 3.03, p = .005) or guilty (M_{moral} = 6.53, 95% CI [6.2, 6.86]; $M_{immoral}$ = 5.85, 95% CI [5.49, 6.21]; b = .68, SE = .26, $t(27.59)$ = 2.66, p = .013). The rest of the results of this model are presented in the main text.

The mixed-effects ANOVA revealed a significant Role x Action Type interaction ($F(1, 343)$ = 31.51, p < .001), such that defense attorneys believed that Harrison's true self was significantly more moral (M = 6.88, 95% CI [6.66, 7.1]) than moral (M = 5.37, 95% CI [5.08, 5.67]; $F(1, 343)$ = 60.84, p < .001). In contrast, prosecuting attorneys did not believe that Harrison's true self was more moral (M = 6.41, 95% CI [6.18, 6.63]) than immoral (M = 6.41, 95% CI [6.16, 6.66]; $F(1, 343)$ = .002, p = .962). The Guilt x Action Type interaction was not significant ($F(1, 343)$ = .35, p = .555). Participants believed that Harrison's true self more moral than immoral irrespective of whether he was described as innocent (M_{moral} = 6.78, 95% CI [6.55, 7.0]; $M_{immoral}$ = 5.96, 95% CI [5.67, 6.25], $F(1, 343)$ = 16.35, p < .001) or guilty (M_{moral} = 6.52, 95% CI [6.3, 6.74]; $M_{immoral}$ = 5.87, 95% CI [5.6, 6.14]; $F(1, 343)$ = 13.26, p < .001).

True self beliefs: Categorical. To test the effect of Role and Action on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for Role (prosecuting attorney vs. defense attorney), Guilt (guilty vs. innocent), Action Type (moral vs. immoral), and their interaction terms, plus random intercepts for participant and vignette. This model was improved by

adding a random slope for Action Type, $\chi^2(2) = 611.26$, $p < .001$, and no additional improvement resulted from adding a second random slope for Role ($\chi^2(2) = 2.59$, $p = .274$) or Guilt ($\chi^2(2) = 1.1$, $p = .576$). Accordingly, the only random slope we retained was the one for Action Type.

Next, we eliminated the fixed effect associated with the three-way interaction. This did not reduce the quality of the fit ($\chi^2(1) = 1.53$, $p = .216$), so we dropped the three-way interaction term from the model. The final model included fixed effects for Role, Guilt, Action Type, and their two-way interaction terms; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction ($b = 2.5$, $SE = .47$, $z = 5.34$, $p < .001$). Defense attorneys were significantly more likely to associate Harrison's true self with moral actions ($M = 77.33\%$, 95% CI [64.28%, 86.61%]) than immoral actions ($M = 35.93\%$, 95% CI [21.41%, 53.59%]; $b = 1.81$, $SE = .5$, $z = 3.62$, $p < .001$). Prosecuting attorneys were no more likely to associate Harrison's true self with moral actions ($M = 64.53\%$, 95% CI [49.28%, 77.31%]) than immoral actions ($M = 78.53\%$, 95% CI [64.09%, 82.43%]; $b = .7$, $SE = .49$, $z = 1.42$, $p = .156$). The Guilt x Action Type interaction was not significant ($b = .12$, $SE = .47$, $z = .25$, $p = .805$). Participants associated Harrison's true self with moral and immoral actions at similar rates irrespective of whether Harrison was described as innocent ($M_{moral} = 78.53\%$, 95% CI [64.09%, 88.23%]; $M_{immoral} = 72.79\%$, 95% CI [58.36%, 83.63%]; $b = .61$, $SE = .51$, $z = 1.21$, $p = .228$) or guilty ($M_{moral} = 69.88\%$, 95% CI [55.51%, 81.18%]; $M_{immoral} = 58.56\%$, 95% CI [41.24%, 74.0%]; $b = .5$, $SE = .48$, $z = 1.03$, $p = .305$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 343) = 23.34$, $p < .001$. Defense attorneys associated Harrison's true self with significantly more moral actions ($M = 3.31$, 95% CI [3.04, 3.58]) than immoral actions ($M = 2.19$, 95% CI [1.89, 2.5]), $F(1, 301) = 29.04$, $p < .001$. Prosecuting attorneys associated Harrisons true self with nonsignificantly fewer moral actions ($M = 2.9$, 95% CI [2.65, 3.15]) than immoral actions ($M = 3.19$, 95% CI [2.91, 3.46]; $F(1, 343) = 1.95$, $p = .163$). The Guilt x Action Type interaction was not significant ($F(1, 343) = .05$, $p = .817$). Participants associated Harrison's true self with more moral than immoral actions irrespective of whether he was described as innocent ($M_{moral} = 3.15$, 95% CI [2.87, 3.43]; $M_{immoral} = 2.71$, 95% CI [2.41, 3.02], $F(1, 343) = 4.05$, $p = .045$) or guilty ($M_{moral} = 3.06$, 95% CI [2.82, 3.31]; $M_{immoral} = 2.69$, 95% CI [2.4, 2.99]; $F(1, 343) = 3.63$, $p = .058$).

Surface self beliefs: Categorical. To test the effect of Role and Action on surface self beliefs, we started with a GLMM that included fixed effects for Role (prosecuting attorney vs. defense attorney), Guilt (guilty vs. innocent), Action Type (moral vs. immoral), and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 513.27$, $p < .001$), and no additional improvement resulted from adding a second random slope for Role ($\chi^2(2) = 2.12$, $p = .347$) or Guilt ($\chi^2(2) = 2.86$, $p = .239$). Accordingly, the only random slope we retained was the one for Action Type.

Next, we eliminated the fixed effect associated with the three-way interaction. This did not reduce the quality of the fit ($\chi^2(1) = .78$, $p = .378$), so we dropped the three-way interaction term from the model. The final model included fixed effects for Role, Guilt, Action Type, and their two-way interaction terms; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction ($b = 1.76$, $SE = .43$, $z = 4.07$, $p < .001$). Defense attorneys were marginally less likely to associate Harrison's surface self with moral actions ($M = 12.16\%$, 95% CI [6.92%, 20.5%]) than immoral actions ($M = 24.89\%$, 95% CI [14.64%, 39.02%]; $b = .87$, $SE = .48$, $z = 1.82$, $p = .069$). In contrast, prosecuting attorneys were marginally more likely to associate Harrison's surface self with moral actions ($M = 25.97\%$, 95% CI [16.13%, 39.02%]) than immoral actions ($M = 12.59\%$, 95% CI [6.92%, 21.8%]; $b = .89$, $SE = .47$, $z = 1.88$, $p = .06$). The Guilt x Action Type interaction was not significant ($b = .09$, $SE = .43$, $z = .21$,

p = .833). Participants associated Harrison's surface self with moral and immoral actions at similar rates irrespective of whether he was described as innocent (M_{moral} = 17.07%, 95% CI [9.94%, 27.74%]; $M_{immoral}$ = 17.59%, 95% CI [9.82%, 29.51%], b = .04, SE = .49, z = .08, p = .94) or guilty (M_{moral} = 19.1%, 95% CI [11.49%, 30.05%]; $M_{immoral}$ = 18.27%, 95% CI [10.49%, 29.88%]; b = .14, SE = .25, z = .55, p = .58).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 343)$ = 15.47, p < .001. Defense attorneys associated Harrison's surface self with significantly fewer moral actions (M = 1.19, 95% CI [.96, 1.42]) than immoral actions (M = 1.85, 95% CI [1.57, 2.13]; $F(1, 343)$ = 11.63, p < .001). Prosecuting attorneys associated Harrison's surface self with significant more moral actions (M = 1.74, 95% CI [1.51, 1.97]) than immoral actions (M = 1.34, 95% CI [1.1, 1.57]; $F(1, 343)$ = 4.56, p = .033). The Guilt x Action Type interaction was not significant ($F(1, 343)$ < .001, p = .992). Participants associated Harrison's surface self with as many moral actions as immoral actions irrespective of whether he was described as innocent (M_{moral} = 1.43, 95% CI [1.18, 1.68]; $M_{immoral}$ = 1.55, 95% CI [1.28, 1.81]; $F(1, 343)$ = .33, p = .566) or guilty (M_{moral} = 1.51, 95% CI [1.29, 1.72]; $M_{immoral}$ = 1.62, 95% CI [1.37, 1.88]; $F(1, 343)$ = .39, p = .531).

Experiment 4

True self beliefs: Continuous. To test the effects of Role (prosecuting attorneys vs. learner), Action Type (moral vs. immoral), and Order on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 1170.6, p < .001), and was further improved by adding a second random slope for Order ($\chi^2(2)$ = 8.75, p = .013). No additional improvement came from adding a random slope for Role ($\chi^2(3)$ = .16, p = .984) or from adding a random slope for the Role x Order interaction ($\chi^2(7)$ = 6.07, p = .532). Accordingly, we the only random slopes retained were the ones for Action Type and Order. Next, we removed the fixed effect for the three-way interaction term. This did not reduce model fit ($\chi^2(1)$ = .39, p = .533), so we dropped the three-way interaction term from the model. We then removed the fixed effects for the two-way interactions. This reduced model fit ($\chi^2(3)$ = 15.27, p = .002), so we retained the two-way interaction terms in the model.

The final model included fixed effects for Role, Action Type, Order, and their two-way interaction terms; random intercepts for participant and vignette and random slopes for Action Type and Order. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 357)$ = 14.52, p < .001. Learners believed that Harrison's true self was significantly more moral (M = 6.8, 95% CI [6.56, 7.04]) than immoral (M = 5.62, 95% CI [5.31, 5.94]), $F(1, 357)$ = 27.27, p < .001. Prosecutors, however, did not believe that Harrison's true self was more moral (M = 6.17, 95% CI [5.86, 6.48]) than immoral (M = 6.25, 95% CI [5.95, 6.54]; $F(1, 357)$ = .10, p = .755).

True self beliefs: Categorical. To test the effects of Role, Action Type, and Order on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 433.16, p < .001), and was improved further by adding a second random slope for Order ($\chi^2(2)$ = 7.54, p = .023). No additional improvement came from adding a random slope for Role ($\chi^2(3)$ = .34, p = .952), or from adding a random slope for the Role x Order interaction ($\chi^2(7)$ = .77, p = .998). Next, we removed the fixed effect for the three-way interaction term. This did not reduce the model's fit ($\chi^2(1)$ = .19, p = .666) so we dropped the three-way interaction term from the model. We then removed the two-way

interaction terms. This significantly reduced the model's fit ($\chi^2(3) = 18.43, p < .001$), so we retained the two-way interaction terms in the model.

The final model included fixed effects for Role, Action Type, Order, and their two-way interaction terms; random intercepts for participant and vignette; and random slopes for Action Type and Order. This model revealed a significant Role x Action Type interaction ($b = 2.19$, $SE = .57$, $z = 3.84$, $p < .001$). Learners were significantly more likely to associate Harrison's true self with moral actions ($M = 68.87\%$, 95% CI [67.65%, 70.02%]) than immoral actions ($M = 46.47\%$, 95% CI [45.05%, 47.84%]; $b = 1.58$, $SE = .47$, $z = 3.32$, $p < .001$). In contrast, prosecutors were not more likely to associate Harrison's true self with moral actions ($M = 52.26\%$, 95% CI [50.98%, 53.56%]) than immoral actions ($M = 60.84\%$, 95% CI [59.45%, 62.14%]; $b = .62$, $SE = .49$, $z = 1.27$, $p = .204$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 357) = 13.05$, $p < .001$. Learners associated Harrison's true self with significantly more moral actions ($M = 1.98$, 95% CI [1.82, 2.15]) than immoral actions ($M = 1.43$, 95% CI [1.26, 1.61]; $F(1, 357) = 14.89$, $p < .001$). Prosecuting attorneys, however, did not associate Harrison's true self with more moral actions ($M = 1.58$, 95% CI [1.4, 1.76]) than immoral actions ($M = 1.78$, 95% CI [1.6, 1.96]; $F(1, 357) = 1.75$, $p = .186$).

Surface self beliefs: Categorical. To test the effects of Role, Action Type, and Order on surface self beliefs, we started with a GLMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 362.15$, $p < .001$), and was improved further by adding a second random slope for Order ($\chi^2(2) = 8.06$, $p = .018$). No additional improvement came from adding a random slope for Role ($\chi^2(3) = .10$, $p = .992$), or from adding a random slope for the Role x Order interaction ($\chi^2(7) = .20$, $p > .999$). Next, we removed the fixed effect for the three-way interaction term. This did not reduce the model's fit ($\chi^2(1) = .03$, $p = .872$) so we dropped the three-way interaction term from the model. We then removed the two-way interaction terms. This significantly reduced the model's fit ($\chi^2(3) = 9.86$, $p = .02$), so we retained the two-way interaction terms in the model.

Our final model included fixed effects for Action Type, Role, Order, and their two-way interaction terms; random intercepts for participant and vignette; and random slopes for Action Type and Order. This model revealed a significant Role x Action Type interaction ($b = 1.58$, $SE = .52$, $z = 3.04$, $p = .002$). Prosecuting attorneys were significantly more likely to associate Harrison's surface self with moral actions ($M = 40\%$, 95% CI [38.91%, 41.13%]) than immoral actions ($M = 27.61\%$, 95% CI [26.44%, 28.85%]; $b = 1.01$, $SE = .43$, $z = 2.34$, $p = .019$). In contrast, learners were not more likely to associate Harrison's surface self with moral actions ($M = 24.27\%$, 95% CI [23.35%, 25.16%]) than immoral actions ($M = 34.6\%$ 95% CI [33.46%, 35.88%]; $b = .57$, $SE = .42$, $z = 1.37$, $p = .171$).

A mixed-effects ANOVA yielded a significant Role x Action Type interaction ($F(1, 357) = 9.01$, $p = .003$). Prosecuting attorneys associated Harrison's surface self with significantly more moral actions ($M = 1.21$, 95% CI [1.03, 1.38]) than immoral actions ($M = .88$, 95% CI [.72, 1.05]; $F(1, 357) = 5.2$, $p = .023$). In contrast, learners associated Harrison's surface self with marginally fewer moral actions ($M = .80$, 95% CI [.66, .94]) than immoral actions ($M = 1.06$, 95% CI [.89, 1.22]; $F(1, 357) = 3.83$, $p = .051$).

Experiment 5

True self beliefs: Continuous. To test the effects of Role (defense attorney vs. learner), Action Type (moral vs. immoral), and Order on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 1407.7$, $p < .001$), and was further improved by adding a second random slope for Order ($\chi^2(2) = 8.24$, $p = .017$). No additional improvement came from adding a random slope for Role ($\chi^2(3) = .09$, $p = .993$) or from adding a random slope for the Role x Order interaction ($\chi^2(7) = 4.44$, $p = .728$). Accordingly, we the only random slopes retained were the ones for Action Type and Order. Next, we removed the fixed effect for the three-way interaction term. This did not reduce model fit ($\chi^2(1) = .43$, $p = .511$), so we dropped the three-way interaction term from the model. We then removed the fixed effects for the two-way interactions. This reduced model fit ($\chi^2(3) = 15.33$, $p = .002$), so we retained the the-way interaction terms in the model.

Our final model included fixed effects for Role, Action Type, Order, and their two-way interaction terms; random intercepts for participant and vignette; and random slopes for Action Type and Order. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 377) = 15.03$, $p < .001$. Defense attorneys believed that Harrison's true self was marginally more moral ($M = 6.38$, 95% CI [6.1, 6.65]) than immoral ($M = 5.94$, 95% CI [5.61, 6.27]; $F(1, 377) = 3.35$, $p = .068$), whereas learners believed that Harrison's true self was significantly less moral ($M = 5.67$, 95% CI [5.39, 5.94]) than immoral ($M = 6.55$, 95% CI [6.27, 6.82]; $F(1, 377) = 13.26$, $p < .001$).

True self beliefs: Categorical. To test the effects of Role, Action Type, and Order on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 706.93$, $p < .001$), and was improved further by adding a second random slope for Order ($\chi^2(2) = 5.25$, $p = .073$). No additional improvement came from adding a random slope for Role ($\chi^2(3) = .55$, $p = .907$), or from adding a random slope for the Role x Order interaction ($\chi^2(7) = 1.43$, $p = .985$). Next, we removed the fixed effect for the three-way interaction term. This did not reduce the model's fit ($\chi^2(1) = .09$, $p = .761$) so we dropped the three-way interaction term from the model. We then removed the two-way interaction terms. This significantly reduced the model's fit ($\chi^2(3) = 8.37$, $p = .039$), so we retained the two-way interaction terms in the model.

Our final model included fixed effects for Role, Action Type, Order and their interaction terms; random intercepts for participant and vignette; and random slopes for Action Type and Order. This model revealed a significant Role x Action Type interaction ($b = 2.06$, $SE = .79$, $z = 2.61$, $p = .009$). Learners were significantly less likely to associate Harrison's true self with moral actions ($M = 42.18\%$, 95% CI [40.84%, 43.52%]) relative to immoral actions ($M = 70.38\%$, 95% CI [68.95%, 71.78%]; $b = 3.15$, $SE = .63$, $z = 5.01$, $p < .001$), whereas defense attorneys were only marginally less likely to associate Harrison's true self with moral actions ($M = 51.65\%$, 95% CI [50.32%, 52.92%]) relative to immoral actions ($M = 58.36\%$, 95% CI [56.81%, 59.98%]; $b = 1.09$, $SE = .60$, $z = 1.84$, $p = .066$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 377) = 5.78$, $p = .017$. Learners associated Harrison's true self with significantly fewer moral actions ($M = 1.3$, 95% CI [1.12, 1.48]) than immoral actions ($M = 2.06$, 95% CI [1.89, 2.23]; $F(1, 377) = 23.28$, $p < .001$), whereas defense attorneys did not associate Harrison's true self with fewer moral actions ($M = 1.52$, 95% CI [1.34, 1.7]) than immoral actions ($M = 1.75$, 95% CI [1.56, 1.93]; $F(1, 377) = 2.15$, $p = .144$).

Surface self beliefs: Categorical. To test the effects of Role, Action Type, and Order on surface self beliefs, we started with a GLMM that included fixed effects for each factor and their interaction terms, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 611.76$, $p < .001$). No further improvement came from adding a second random slope for Order ($\chi^2(2) = 4.25$, $p = .12$) or Role ($\chi^2(2) = 2.3$, $p = .317$), so these terms were left out. Next, we removed the fixed effect for the three-way interaction term. This did not reduce the quality of the model's fit ($\chi^2(1) = .24$, $p = .628$) so we dropped the three-way interaction term from the model. We then removed the two-way interaction terms. This reduced the model's fit to a degree that approached significance ($\chi^2(3) = 7.39$, $p = .06$), so we retained the two-way interaction terms.

The final model included fixed effects for Action Type, Role, Order, and their two-way interactions; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction ($b = 1.76$, $SE = .73$, $z = 2.43$, $p = .015$). Learners were significantly likelier to associate Harrison's surface self with moral actions ($M = 54.06\%$, 95% CI [52.75%, 55.4%]) relative to immoral actions ($M = 21.47\%$ 95% CI [20.21%, 22.74%]; $b = 3.59$, $SE = .62$, $z = 5.78$, $p < .001$). Defense attorneys had a weaker, but still significant tendency to associate Harrison's surface self with moral actions ($M = 43.8\%$, 95% CI [42.56%, 45.04%]) more often than immoral actions ($M = 30.06\%$, 95% CI [28.65%, 31.47%]; $b = 1.83$, $SE = .57$, $z = 3.21$, $p = .001$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 377) = 5.15$, $p = .024$. Learners associated Harrison's surface self significantly with more moral actions ($M = 1.59$, 95% CI [1.41, 1.77]) than immoral actions ($M = .71$, 95% CI [.56, .87]; $F(1, 377) = 34.92$, $p < .001$. Defense attorneys had a weaker, but still significant tendency to associate Harrison's surface self with more moral actions ($M = 1.34$, 95% CI [1.17, 1.52]) than immoral actions ($M = .94$, 95% CI [.77, 1.11]; $F(1, 377) = 7.6$, $p = .006$).

Experiment 6

True self beliefs: Continuous. To test the effect of Role (campaign manager vs. candidate) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus a random intercept for participant. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 449.73$, $p < .001$), so we retained it in the final model.

The final model revealed a significant Role x Action Type interaction ($b = 1.56$, $SE = .27$, $t(296) = 5.87$, $p < .001$). Campaign managers believed that dm88's true self was significantly more moral ($M = 6.28$, 95% CI [6.19, 6.37]) than immoral ($M = 4.48$, 95% CI [4.38, 4.58]; $b = 1.8$, $SE = .19$, $t(296) = 9.6$, $p < .001$). In contrast, candidates did not believe that dm88's true self was more moral ($M = 5.9$, 95% CI [5.8, 5.99]) than immoral ($M = 5.66$, 95% CI [5.56, 5.75]; $b = .24$, $SE = .19$, $t(296) = 1.27$, $p = .205$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 296) = 34.41$, $p < .001$. Campaign managers believed that dm88's true self with significantly more moral ($M = 6.3$, 95% CI [6.01, 6.6]) than immoral ($M = 4.49$, 95% CI [4.21, 4.78]), $F(1, 296) = 92.23$, $p < .001$, whereas candidates did not believe that dm88's true self was more moral ($M = 5.9$, 95% CI [5.61, 6.18]) than immoral ($M = 5.66$, 95% CI [5.36, 5.95]; $F(1, 296) = 1.62$, $p = .205$).

True self beliefs: Categorical. To test the effects of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for both factors and

their interaction term, plus a random intercept for participant. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 309.96$, $p < .001$), so we retained it in the model.

The final model revealed a marginal Role x Action Type interaction ($b = 2.34$, $SE = 1.29$, $z = 1.81$, $p = .07$). Campaign managers were significantly likelier to associate dm88's true self with moral actions ($M = 59.18\%$, 95% CI [56.56%, 61.81%]) than with immoral actions ($M = 27.26\%$, 95% CI [25.04%, 29.61%]; $b = 17.77$, $SE = 1.26$, $z = 14.11$, $p < .001$). Candidates had a weaker, but still significant tendency to associate dm88's true self with moral actions ($M = 56.73\%$, 95% CI [53.93%, 59.42%]) more often than with immoral actions ($M = 30.89\%$, 95% CI [28.43%, 33.24%]; $b = 15.43$, $SE = 1.34$, $z = 11.55$, $p < .001$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 296) = 18.63$, $p < .001$. Campaign managers associated dm88's true self with significantly more moral actions ($M = 1.22$, 95% CI [1.07, 1.37]) than immoral actions ($M = .39$, 95% CI [.27, .51]; $F(1, 296) = 71.95$, $p < .001$). Candidates had a weaker, but still significant tendency to associate dm88's true self with more moral actions ($M = 1.05$, 95% CI [.90, 1.21]) than immoral actions ($M = .82$, 95% CI [.67, .98]; $F(1, 296) = 5.48$, $p = .02$).

Surface self beliefs: Categorical. To test the effects of Role and Action Type on surface self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus a random intercept for participant. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 236.12$, $p < .001$), which we retained in the final model.

The final model revealed a non-significant Role x Action Type interaction ($b = 1.81$, $SE = 1.10$, $z = 1.64$, $p = .10$). Campaign managers were significantly likelier to associate dm88's surface self with immoral actions ($M = 59.66\%$, 95% CI [57.5%, 62.1%]) relative to moral actions ($M = 27.92\%$, 95% CI [25.73%, 30.35%]; $b = 9.89$, $SE = .96$, $z = 10.35$, $p < .001$). Candidates also were significantly likelier to associate dm88's surface self with immoral actions ($M = 46.52\%$, 95% CI [44.26%, 48.77%]) relative to moral actions ($M = 28.99\%$, 95% CI [26.66%, 31.47%]; $b = 8.08$, $SE = .93$, $z = 8.68$, $p < .001$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 296) = 4.5$, $p = .035$. Campaign managers associated dm88's surface self with significantly more immoral actions ($M = 1.13$, 95% CI [.98, 1.28]) than moral actions ($M = .53$, 95% CI [.39, .66]), $F(1, 296) = 39.01$, $p < .001$. Candidates had a weaker, but still significant tendency to associate dm88's surface self with more immoral actions ($M = .96$, 95% CI [.81, 1.11]) than moral actions ($M = .65$, 95% CI [.51, .79]; $F(1, 296) = 10.33$, $p = .001$).

Behavior. To test the effects of Role and Action Type on behavior, we started with an LMM that included fixed effects for both factors and their interaction term, plus a random intercept for participant. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 566.83$, $p < .001$), so we retained it in our final model.

The final model revealed a significant Role x Action Type interaction ($b = 1.77$, $SE = .26$, $t(296.02) = 6.71$, $p < .001$). Campaign managers rewarded dm88's moral actions ($M = 4.14$, 95% CI [4.05, 4.24]) significantly more than they punished dm88's immoral actions ($M = 3.02$, 95% CI [2.93, 3.10]; $b = 1.24$, $SE = .19$, $t(296.02) = 6.63$, $p < .001$). Conversely, candidates punished dm88's immoral actions ($M = 4.16$, 95% CI [4.08, 4.25]) significantly more than they rewarded dm88's moral actions ($M = 3.60$, 95% CI [3.50, 3.69]; $b = .54$, $SE = .19$, $t(296.02) = 2.86$, $p = .005$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 296) = 44.97$, $p < .001$. Campaign managers rewarded dm88's moral actions ($M = 4.27$, 95% CI [4.06, 4.49]) significantly more than they punished dm88's immoral actions ($M = 3.04$, 95% CI [2.84, 3.23]; $F(1, 296) = 44.01$, $p < .001$). Conversely, candidates punished dm88's immoral actions ($M = 4.18$, 95%

CI [3.95, 4.4]) significantly more than they rewarded dm88's moral actions (M = 3.64, 95% CI [3.42, 3.86]; $F(1, 296) = 8.19$, $p = .005$).

Experiment 7

True self beliefs: Continuous. To test the effects of Role (prosecuting vs. defense attorney) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 1264.4$, $p < .001$), and was further improved by adding a second random slope for Role ($\chi^2(2) = 7.68$, $p = .022$). Accordingly, we retained both random slopes in the model.

The final model included fixed effects for Role, Action Type, and their interaction; random intercepts for participant and vignette; and random slopes for Action Type and Role. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 325) = 43.46$, $p < .001$. Defense attorneys believed that Harrison's true self with significantly more moral (M = 7.02, 95% CI [6.8, 7.23]) than immoral (M = 6.16, 95% CI [5.86, 6.45]; $F(1, 325) = 19.14$, $p < .001$). Conversely, prosecuting attorneys believed that Harrison's true self was more immoral (M = 6.95, 95% CI [6.74, 7.16]) than moral (M = 5.96, 95% CI [5.69, 6.22]; $F(1, 325) = 24.43$, $p < .001$).

True self beliefs: Categorical. To test the effects of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 517.08$, $p < .001$), and was not improved further by adding a second random slope for Order ($\chi^2(2) = 1.19$, $p = .551$). Accordingly, the only random slope we retained in the model was the one for Action Type.

Our final model included; fixed effects for Role, Action Type, and their interaction term; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction ($b = 2.55$, $SE = .43$, $z = 5.94$, $p < .001$). Prosecuting attorneys were significantly likelier to associate Harrison's true self with immoral actions (M = 78.51%, 95% CI [77.69%, 79.24%]) relative to moral actions (M = 47.21%, 95% CI [46.18%, 48.21%]; $b = 2.05$, $SE = .58$, $z = 3.54$, $p < .001$). In contrast, defense attorneys associated Harrison's true self with moral actions (M = 67.95%, 95% CI [67.12%, 68.88%]) and immoral actions (M = 60.6%, 95% CI [59.63%, 61.57%]) at similar rates ($b = .5$, $SE = .57$, $z = .87$, $p = .384$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 325) = 34.23$, $p < .001$. Prosecuting attorneys associated Harrison's true self with significantly more immoral actions (M = 3.88, 95% CI [3.67, 4.09]) than moral actions (M = 2.47, 95% CI [2.21, 2.73]; $F(1, 325) = 45.63$, $p < .001$). In contrast, defense attorneys associated Harrison's true self with about as many moral actions (M = 3.35, 95% CI [3.1, 3.59]) as immoral actions (M = 3.05, 95% CI [2.79, 3.31]; $F(1, 325) = 2.14$, $p = .144$).

Surface self beliefs: Categorical. To test effects of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 460.45$, $p < .001$), and was not improved further by adding a second random slope for Order ($\chi^2(2) = 1.06$, $p = .588$). Accordingly, the only random slope we retained was the one for Action Type.

Our final model included fixed effects for Action Type, Role, and their interaction term; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction (b = 2.33, SE = .42, z = 5.52, p < .001). Prosecuting attorneys were significantly likelier to associate Harrison's surface self with moral actions (M = 47.43%, 95% CI [46.45%, 48.36%]) than with immoral actions (M = 16.65%, 95% CI [16.03%, 17.31%]; b = 2.1, SE = .61, z = 3.46, p < .001). In contrast, defense attorneys associated Harrison's surface self with moral actions (M = 29.25%, 95% CI [28.32%, 30.15%]) and immoral actions (M = 31.82% 95% CI [30.96%, 32.66%]) at similar rates (b = .23, SE = .60, z = .38, p = .706).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 325)$ = 28.98, p < .001. Prosecuting attorneys associated Harrison's surface self with more moral actions (M = 2.26, 95% CI [2.01, 2.51]) than immoral actions (M = .88, 95% CI [.69, 1.06]; $F(1, 325)$ = 49.65, p < .001), whereas defense attorneys associated Harrison's surface self with about as many moral actions (M = 1.5, 95% CI [1.26, 1.73]) as immoral actions (M = 1.59, 95% CI [1.36, 1.83]; $F(1, 325)$ = .25, p = .619).

Experiment 8

True self beliefs: Continuous. To test the effects of Role (prosecuting vs. defense attorney) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 1416.9, p < .001), but was not improved further by adding a second random slope for Role ($\chi^2(2)$ = 1.08, p = .582). Accordingly, the only random slope we retained was the one for Action Type.

The final model included fixed effects for Role, Action Type, and their interaction; random intercepts for participant and vignette; and a random slope for Action Type. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 299)$ = 35.32, p < .001. Defense attorneys believed that Harrison's true self with significantly more moral (M = 7.44, 95% CI [7.24, 7.63]) than immoral (M = 5.98, 95% CI [5.7, 6.27]; $F(1, 299)$ = 45.99, p < .001). Conversely, prosecuting attorneys believed that Harrison's true self was marginally more immoral (M = 6.47, 95% CI [6.17, 6.77]) than moral (M = 6.1, 95% CI [5.8, 6.4]; $F(1, 299)$ = 2.78, p = .097).

True self beliefs: Categorical. To test the effects of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 518.22, p < .001), and was not improved further by adding a second random slope for Role ($\chi^2(2)$ = .01, p = .996). Accordingly, the only random slope we retained in the model was the one for Action Type.

Our final model included; fixed effects for Role, Action Type, and their interaction term; random intercepts for participant and vignette; and a random slope for Action Type. This model revealed a significant Role x Action Type interaction (b = 2.67, SE = .5, z = 5.38, p < .001). Prosecuting attorneys associated Harrison's true self with moral actions (M = 65.83%, 95% CI [50.19%, 78.65%]) and immoral actions (M = 71.69%, 95% CI [56.45%, 83.19%]) at similar rates (b = .27, SE = .5, z = .54, p = .588). In contrast, defense attorneys were significantly more likely to

associate Harrison's true self with moral actions (M = 92.08%, 95% CI [85.2%, 95.92%]) than immoral actions (M = 51.33%, 95% CI [35.46%, 66.94%]; b = 2.4, SE = .52, z = 4.64, p < .001).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 299)$ = 26.85, p < .001. Prosecuting attorneys associated Harrison's true self with as many moral actions (M = 2.96, 95% CI [2.68, 3.24]) as immoral actions (M = 3.09, 95% CI [2.79, 3.38]; $F(1, 299)$ = .34, p = .563). In contrast, defense attorneys associated Harrison's true self with significantly more moral actions (M = 4.03, 95% CI [3.82, 4.25]) than immoral actions (M = 2.55, 95% CI [2.27, 2.83]; $F(1, 299)$ = 46.26, p < .001).

Surface self beliefs: Categorical. To test effects of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 451.05, p < .001), and was not improved further by adding a second random slope for Role ($\chi^2(2)$ =.19, p = .909). Accordingly, the only random slope we retained in the model was the one for Action Type.

Our final model included fixed effects for Action Type, Role, and their interaction term; random intercepts for participant and vignette; and random slopes for Action Type and role. This model revealed a significant Role x Action Type interaction (b = 2.19, SE = .46, z = 4.78, p < .001). Prosecuting attorneys associated Harrison's surface self with moral actions (M = 20.39%, 95% CI [11.79%, 32.92%]) and immoral actions (M = 17.75%, 95% CI [10.16%, 29.18%]) at similar rates (b = .17, SE = .5, z = .34, p = .732). In contrast, defense attorneys were significantly less likely to associate Harrison's surface self with moral actions (M = 6.87%, 95% CI [3.62%, 12.65%]) than with immoral actions (M = 35.65% 95% CI [22.77%, 50.99%]; b = 2.02, SE = .50, z = 4.0, p < .001).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 299)$ = 21.86, p < .001. Prosecuting attorneys associated Harrison's surface self with as many moral actions (M = 1.55, 95% CI [1.29, 1.82]) as immoral actions (M = 1.42, 95% CI [1.16, 1.68]); $F(1, 299)$ = .42, p = .516), whereas defense attorneys associated Harrison's surface self with significantly fewer moral actions (M = .82, 95% CI [.63, 1.01]) than immoral actions (M = 2.05, 95% CI [1.79, 2.31]; $F(1, 299)$ = 36.08, p < .001).

Experiment 9

True self beliefs: Continuous. To test the effects of Role (Harrison's attorney vs. Ineos' attorney) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2)$ = 689.01, p < .001), and was further improved by adding a second random slope for Role ($\chi^2(2)$ = 11.89, p = .003). Accordingly, we retained both random slopes in the model.

The final model included fixed effects for Role, Action Type, and their two-way interaction; random intercepts for participant and vignette; and random slopes for Action Type and Role. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 198)$ = 22.59, p < .001. Lawyers who were to advocate for Harrison believed that Harrison's true self was significantly more moral (M = 6.93, 95% CI [6.68, 7.19]) than immoral (M = 5.93, 95% CI [5.46, 6.2]; $F(1, 198)$ = 24.05, p < .001). In contrast, lawyers who were to advocate for Ineos believed that Harrison's true self was marginally less moral (M = 6.17, 95% CI [5.85, 6.5]) than immoral (M = 6.64, 95% CI [6.31, 6.97]; $F(1, 198)$ = 3.71, p = .056).

True self beliefs: Categorical. To test the effect of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 272.95$, $p < .001$), and was not improved further by adding a second random slope for Order ($\chi^2(2) = .97$, $p = .615$). Accordingly, the only random slope we retained in the model was the one for Action Type.

The final model revealed a significant Role x Action Type interaction ($b = 2.48$, $SE = .53$, $z = 4.66$, $p < .001$). Lawyers who were to advocate for Harrison were significantly likelier to associate Harrison's true self with moral ($M = 68.98\%$, 95% CI [67.74%, 70.16%]) versus immoral actions ($M = 46.8\%$, 95% CI [45.38%, 48.15%]; $b = 1.43$, $SE = .56$, $z = 2.54$, $p = .011$). In contrast, lawyers who were to advocate for Ineos were marginally less likely to associate Harrison's true self with moral ($M = 48.83\%$, 95% CI [47.51%, 50.15%]) versus immoral actions ($M = 64.2\%$, 95% CI [62.96%, 65.49%]; $b = 1.05$, $SE = .58$, $z = 1.83$, $p = .067$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 198) = 19.92$, $p < .001$. Lawyers who were to advocate for Harrison associated Harrison's true self with significantly more moral actions ($M = 3.36$, 95% CI [3.03, 3.69]) than immoral actions ($M = 2.42$, 95% CI [5.46, 6.2]; $F(1, 198) = 14.52$, $p < .001$). Conversely, lawyers who were to advocate for Ineos associated Harrison's true self with significantly fewer moral actions ($M = 2.49$, 95% CI [2.14, 2.85]) than immoral actions ($M = 3.17$, 95% CI [2.84, 3.5]; $F(1, 198) = 6.5$, $p = .012$).

Surface self beliefs: Categorical. To test the effects of Role and Action Type on surface self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 205.46$, $p < .001$), and was not improved further by adding a second random slope for Order ($\chi^2(2) = .09$, $p = .957$). Accordingly, the only random slope we retained was the one for Action Type.

The final model revealed a significant Role x Action Type interaction ($b = 1.78$, $SE = .49$, $z = 3.61$, $p < .001$). Lawyers who were to advocate for Harrison associated Harrison's surface self with moral actions ($M = 22.65\%$, 95% CI [21.72%, 23.61%]) and immoral actions ($M = 33.05\%$, 95% CI [31.88%, 34.23%]) at similar rates ($b = .67$, $SE = .51$, $z = 1.32$, $p = .188$). In contrast, lawyers who were to advocate for Ineos were significantly more likely to associate Harrison's surface self with moral actions ($M = 38.81\%$, 95% CI [37.71%, 39.99%]) than immoral actions ($M = 23.51\%$, 95% CI [22.56%, 24.57%]; $b = 1.11$, $SE = .53$, $z = 2.10$, $p = .035$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 198) = 11.45$, $p < .001$. Lawyers who were to advocate for Harrison associated Harrison's surface self with significantly fewer moral actions ($M = 1.17$, 95% CI [.89, 1.44]) than immoral actions ($M = 1.64$, 95% CI [1.34, 1.95] ; $F(1, 198) = 4.53$, $p = .035$). Conversely, lawyers who were to advocate for Ineos associated Harrison's surface self with significantly more moral actions ($M = 1.88$, 95% CI [1.54, 2.22]) than immoral actions ($M = 1.25$, 95% CI [.94, 1.55]; $F(1, 198) = 6.98$, $p = .009$).

Experiment 10

True self beliefs: Continuous. To test the effects of Role (Harrison's attorney vs. Ineos' attorney) and Action Type (moral vs. immoral) on the continuous measure of true self beliefs, we started with an LMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 552.19$, $p < .001$), but was not improved further by adding a second random

slope for Role ($\chi^2(2) = .004$, $p = .948$). Accordingly, the only random slope we retained was the one for Action Type.

The final model included fixed effects for Role, Action Type, and their two-way interaction; random intercepts for participant and vignette; and a random slope for Action Type. The results of this model are presented in the main text.

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 198) = 8.94$, $p = .003$. Lawyers who were to advocate for Harrison believed that Harrison's true self was nonsignificantly more moral ($M = 6.47$, 95% CI [6.17, 6.77]) than immoral ($M = 6.18$, 95% CI [5.78, 6.58]; $F(1, 198) = 1.78$, $p = .184$). In contrast, lawyers who were to advocate for Ineos believed that Harrison's true self was significantly less moral ($M = 5.99$, 95% CI [5.71, 6.28]) than immoral ($M = 6.59$, 95% CI [6.3, 6.89]; $F(1, 198) = 8.62$, $p = .004$).

True self beliefs: Categorical. To test the effect of Role and Action Type on the categorical measure of true self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 268.95$, $p < .001$), and was not improved further by adding a second random slope for Role ($\chi^2(2) = .05$, $p = .977$). Accordingly, the only random slope we retained in the model was the one for Action Type.

The final model revealed a significant Role x Action Type interaction ($b = 1.06$, $SE = .55$, $z = 2.02$, $p = .044$). Lawyers who were to advocate for Harrison associated Harrison's true self with moral ($M = 65.2\%$, 95% CI [48.54%, 78.82%]) and immoral actions ($M = 71.92\%$, 95% CI [53.77%, 64.94%]) at similar rates ($b = .31$, $SE = .54$, $z = .58$, $p = .561$). In contrast, lawyers who were to advocate for Ineos were significantly less likely to associate Harrison's true self with moral ($M = 45.22\%$, 95% CI [29.64%, 61.8%]) versus immoral actions ($M = 76.45\%$, 95% CI [60.02%, 87.53%]; $b = 1.37$, $SE = .53$, $z = 2.6$, $p = .009$).

A mixed-effects ANOVA revealed a significant Role x Action Type interaction, $F(1, 198) = 3.9$, $p = .05$. Lawyers who were to advocate for Harrison associated Harrison's true self with as many moral actions ($M = 2.97$, 95% CI [2.65, 3.29]) as immoral actions ($M = 3.11$, 95% CI [2.72, 3.49]; $F(1, 198) = .28$, $p = .60$). Conversely, lawyers who were to advocate for Ineos associated Harrison's true self with significantly fewer moral actions ($M = 2.36$, 95% CI [2.01, 2.71]) than immoral actions ($M = 3.21$, 95% CI [2.87, 3.55]; $F(1, 198) = 11.68$, $p < .001$).

Surface self beliefs: Categorical. To test the effects of Role and Action Type on surface self beliefs, we started with a GLMM that included fixed effects for both factors and their interaction term, plus random intercepts for participant and vignette. This model was improved by adding a random slope for Action Type ($\chi^2(2) = 214.67$, $p < .001$), and was not improved further by adding a second random slope for Order ($\chi^2(2) = .09$, $p = .958$). Accordingly, the only random slope we retained was the one for Action Type.

In this model, the Role x Action Type interaction was not significant ($b = .52$, $SE = .55$, $z = .93$, $p = .931$). Lawyers who were to advocate for Harrison were significantly more likely to associate Harrison's surface self with moral actions ($M = 21.1\%$, 95% CI [12.15%, 34.35%]) than immoral actions ($M = 7.27\%$, 95% CI [3.25%, 15.46%]) at similar rates ($b = 1.23$, $SE = .56$, $z = 2.18$, $p = .029$). Likewise, lawyers who were to advocate for Ineos were significantly more likely to associate Harrison's surface self with moral actions ($M = 32.39\%$, 95% CI [20.18%, 47.59%]) than immoral actions ($M = 7.69\%$, 95% CI [3.55%, 15.89%]; $b = 1.75$, $SE = .54$, $z = 3.22$, $p = .001$).

A mixed-effects ANOVA revealed a nonsignificant Role x Action Type interaction, $F(1, 198) = 1.52$, $p = .219$. Lawyers who were to advocate for Harrison associated Harrison's surface self with significantly more moral actions ($M = 1.52$, 95% CI [1.24, 1.81]) than immoral actions ($M =$

1.02, 95% CI [.73, 1.32]; $F(1, 198) = 4.63$, $p = .033$). Likewise, lawyers who were to advocate for Ineos associated Harrison's surface self with significantly more moral actions ($M = 1.94$, 95% CI [1.6, 2.28]) than immoral actions ($M = 1.04$, 95% CI [.77, 1.3]; $F(1, 198) = 16.41$, $p < .001$).

References

1 Westfall, J., Kenny, D. A. & Judd, C. M. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General* **143** (2014).

Supplementary Tables

Supplementary Table 1. Areas of law practiced by sample in Experiment 9

Area	N	Area	N
Public Interest	14	Public Defender	8
Government	22	Private Defense Attorney	30
Private Sector	54	Prosecutor	10
Trial (Civil)	56	Entertainment	3
Trial (Criminal)	19	Real Estate	22
Immigration	3	Digital Media & Internet	4
Estate Planning	23	Legal Malpractice	2
Personal Injury	18	Legal Scholar / Academic	2
Toxic Tort	1	Other	46
Civil Rights	3	Law Degree, Not Practicing	5

Note. Lawyers were permitted to indicate that they practice multiple areas of law. Thus, the total number of lawyers with experience in each area exceed the sample size of 200.

Supplementary Table 2. Areas of law practiced by sample in Experiment 10

Area	N
Litigation	117
Transaction	23
Trusts & Estates	7
Tax & Bankruptcy	11
Other	42

Note. Lawyers were permitted to select only one area of specialization.