

**Title:** The Actor's Insight: Actors have comparable interoception but better metacognition than non-actors

**Running Head:** Interoception and metacognition in actors and non-actors

**Authors:** Sokol-Hessner, Peter<sup>1\*</sup>

Wing-Davey, Mark<sup>2</sup>

Illingworth, Scott<sup>2</sup>

Fleming, Stephen M.<sup>3,4,5</sup>

Phelps, Elizabeth A.<sup>6</sup>

1. Department of Psychology, University of Denver, Denver, CO, USA.

2. Tisch School of The Arts Graduate Acting Program, New York University, New York, NY, USA.

3. Wellcome Centre for Human Neuroimaging, University College London, London, UK.

4. Department of Experimental Psychology, University College London, London, UK.

5. Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, UK

6. Department of Psychology, Harvard University, Cambridge, MA, USA.

\* Corresponding author: [Peter.Sokol-Hessner@du.edu](mailto:Peter.Sokol-Hessner@du.edu)

Data and model files available on the Open Science Framework (OSF) at <https://osf.io/84cpg/>

© 2022, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission.

The final article will be available, upon publication, via its DOI: [10.1037/emo0001080](https://doi.org/10.1037/emo0001080)

**Abstract:** Both accurately sensing our own bodily signals and knowing whether we have accurately sensed them may contribute to a successful emotional life, but there is little evidence on whether these physiological perceptual and metacognitive abilities systematically differ between people. Here, we examined whether actors, who receive substantial training in the production, awareness, and control of emotion, and non-actor controls differed in interoceptive ability (the perception of internal bodily signals) and/or metacognition about interoceptive accuracy (awareness of that perception), and explored potential sources of individual differences in and consequences of these abilities including correlational relationships with state and trait anxiety, proxies for acting ability, and the amount of acting training. Participants performed a heartbeat detection task in which they judged whether tones were played synchronously or delayed relative to their heartbeats, and then rated their metacognitive confidence in that judgment. Cardiac interoceptive accuracy and metacognitive awareness of interoceptive accuracy were independent, and while actors' and controls' interoceptive accuracy was not significantly different, actors had consistently superior metacognitive awareness of interoception. Exploratory analyses additionally suggest that this metacognitive ability may be correlated with measures of acting ability, but not the duration of acting training. Interoceptive accuracy and metacognitive insight into that accuracy appear to be separate abilities, and while actors may be no more accurate in reading their bodies, their metacognitive insight means they know better when they're accurate and when they're not.

**Keywords:**

Interoception; Metacognition; Acting; Emotion; Training

## **I. Introduction**

Successfully navigating our emotional lives is critical to adaptive functioning and mental health (Khalsa et al., 2018). Because emotion itself consists of multiple coordinated components (including physiological responses) (Scherer, 2005), multiple different monitoring and regulation processes are involved in this navigation. One such process, the perception of internal bodily events (called interoceptive sensitivity or accuracy) has been the focus of much recent research, as well as a related concept, the meta-knowledge of how accurate our bodily perceptions are (metacognitive awareness of interoceptive accuracy, known as metacognitive efficiency). Understanding individual differences and potential correlates of these two abilities is thus of critical importance. Here we used a cardiac interoceptive task to assess both interoceptive accuracy and metacognitive efficiency, and tested for potential differences in either ability in a unique group of participants: actors. Both interoceptive accuracy and awareness of that accuracy could be extremely important in acting, which as a profession requires the production, awareness, communication, and control of physiological and bodily responses, all of which would benefit from a greater ability to perceive one's own physiology, and know the accuracy of one's perception at that moment in time.

Interoceptive accuracy, defined as the extent to which one makes correct interoceptive judgments, contributes to individual differences in the role of emotion in a variety of domains. Better interoceptors report more intense emotional experiences (Barrett et al., 2004; Wiens et al., 2000), perhaps especially in the negative domain (Critchley et al., 2004), experience greater empathy for others (Ernst et al., 2013; Heydrich et al., 2021; Terasawa et al., 2014), and use emotion regulation more often (Kever et al., 2015). Similarly, interoceptive accuracy is relevant

to issues of mental health (Khalsa et al., 2018), including alcohol (Ateş Çöl et al., 2016) and drug dependence (Gray & Critchley, 2007; Stewart, Juavinett, et al., 2015; Stewart, May, et al., 2015) and may contribute to autism spectrum disorder (Garfinkel et al., 2016) and other psychiatric disorders (Bonaz et al., 2021). Interoceptive accuracy has also been linked to decision-making in settings in which emotions are thought to play a critical role, like fast and complicated high-stakes choices (Kandasamy et al., 2016) and loss aversion (Sokol-Hessner et al., 2015). These links suggest that interoceptive accuracy is one contributing factor to the intensity of the felt emotional experience, and insofar as felt emotions contribute to expressed and thus perceived emotions (Gosselin et al., 1995), may therefore alter the communication of emotion more broadly.

The metacognitive awareness of one's own interoceptive accuracy is a distinct but related construct within interoception, referring to the ability to be aware of or sensitive to one's own interoceptive performance (Fleming & Lau, 2014). Metacognitive awareness or efficiency (here used interchangeably) is often measured with confidence ratings collected alongside task judgments, with good metacognition corresponding to a close relationship between judgment accuracy and confidence, on a per-judgment basis. For example, two people's confidence judgments over the course of the experiment may have the same average and variance, but two radically different levels of metacognitive sensitivity. In a person with high levels of metacognitive sensitivity, trials with high confidence would be very likely to be correct trials, while trials with low confidence would be at chance levels of performance. This person's confidence would closely track their accuracy, on a trial-by-trial basis. In contrast, a person with very low levels of metacognitive sensitivity would be no more likely to be correct on trials when

their confidence judgments were high than those when their confidence judgments were low. Good metacognitive sensitivity can therefore be thought of as resulting from a close *correspondence*, trial-by-trial, between accuracy and confidence.

The links between interoception and metacognition have only recently begun to be examined, perhaps because of statistical challenges inherent in their measurement, including bias, robustness, and the computational sophistication required to accurately estimate metacognitive abilities (Brener & Ring, 2016; Fleming, 2017; Fleming & Lau, 2014). Early studies have found that metacognitive awareness of interoception is related to metacognitive awareness in other, non-emotional domains (Chua & Bliss-Moreau, 2016), dissociable from interoceptive accuracy (Garfinkel et al., 2014; Harrison et al., 2021; but see Forkmann et al., 2016), and may mitigate impairments in recognizing prosody in people with autism spectrum disorder (Mulcahy et al., 2019). These findings indicate that metacognitive awareness of interoceptive accuracy is an important and independent construct within interoception, but we do not yet know the full extent to which internal bodily perception and interoceptive metacognitive efficiency may relate to one another, have real-world consequences, or be altered by years of training.

There are significant professional incentives for actors to be better at both interoception and metacognition compared to non-actors. While many decades-old popular approaches to acting training emphasize bodily awareness (Stanislavsky, 1952; Strasberg, 1987), there is little evidence on whether any quantitative emotional differences exist between actors and non-actors (but see Noice & Noice, 2006; Goldstein & Bloom, 2011), let alone evidence on the source of those differences (e.g. self-selection into acting professions, effects of advanced training). What

evidence there is suggests that actors may be different, especially in terms of emotion-related processes and traits, than non-actors. For example, evidence indicates that actors are more open, assertive, creative, and volatile than non-actors (Dumas et al., 2020), that felt emotions are crucial to actors' accurate portrayal of emotions (Gosselin et al., 1995), that acting training supports children's emotional development (Goldstein & Lerner, 2018), and that self-reported interoceptive ability interacts with types of acting training to predict emotion simulation success (Jackson & Muir, 2019). Improved characterization of any differences between actors and non-actors, and the source of those differences, could have significant consequences for mental health across clinical and non-clinical populations, especially if there is evidence that training as an actor produces these differences, a possibility made more likely by recent findings that training may improve some metacognitive abilities (Carpenter et al., 2019). Using actors as a test population, therefore, we can ask to what extent significant objective, quantitative differences even exist in the accurate perception of physiological signals and/or the metacognitive awareness of the quality of that perception, and whether there is any evidence that the duration of acting training might have influenced these abilities.

In the current study, actors and non-actor controls performed a heartbeat-detection task to quantify interoceptive sensitivity and metacognitive sensitivity to trial-wise interoceptive accuracy, in a robust and novel design. Metacognitive awareness (meta- $d'/d'$ ) was analyzed in a powerful, hierarchical Bayesian framework that both pooled data and allowed for individual differences (Fleming, 2017). We related these measures to individual differences in self-report measures of task experience and anxiety, as well as measures of acting ability and acting training duration to explore variability within participants, and between actors and non-actors. As acting

may involve the production, awareness, control, and expression of emotion, including bodily signals (e.g. breathing, heart rate, tenseness, etc), we hypothesized that actors would show enhanced interoceptive accuracy and enhanced metacognitive efficiency to interoception, compared to non-actors. As much is not yet known about the intersections between interoceptive accuracy, metacognition about interoceptive accuracy, individual differences in these processes, and acting (including being an actor, ability in acting, or acting training), we conducted additional exploratory analyses of these factors.

## **II. Methods**

The study was reviewed and approved by the New York University institutional review board (the University Committee on Activities Involving Human Subjects). Below, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. This study was not preregistered.

### *Participants*

Participants were recruited via emails sent to current and former students of New York University's Tisch School of Graduate Acting Program, and general population advertisements using flyers and online participant recruitment pools. The sample size was determined and limited by our ability to recruit the unique participants in this study (see below). Recruitment proceeded until we could recruit no more Tisch Graduate Actors to participate, at which point recruitment ended and analysis began. Controls were recruited in parallel to roughly match the number of Actors.

Twenty-eight current or former students from New York University's Tisch School of the Arts Graduate Acting Program were recruited. Two were dropped from analysis (one for watching his heartbeat move his glasses and clothes; one for failing to understand the task), resulting in a final N of 26 Tisch Graduate Acting students or alumni (12 females; mean (standard deviation) age = 27.2 years (3.0 years), BMI = 22.0 kg/m<sup>2</sup> (2.6 kg/m<sup>2</sup>)).

Thirteen controls reported significant acting experience, and so were categorized as "Other" actors for the purposes of analysis. None were current or former Tisch students. Two were dropped (one because of technical difficulty with electrode adherence; one for having a sufficiently high resting heartrate that the delayed tone playback was significantly affected), resulting in a final N of 11 "Other" actors (9 females; mean (standard deviation) age = 22.0 years (3.1 years), BMI = 23.1 kg/m<sup>2</sup> (4.0 kg/m<sup>2</sup>)).

Thirty-two non-actor control participants were recruited. Three were dropped (two for failing to understand the task; one for having a sufficiently high resting heartrate that delayed-tone playback was significantly affected), resulting in a final N of 29 non-actor controls (23 females; mean (standard deviation) age = 24.2 years (5.1 years), BMI = 22.9 kg/m<sup>2</sup> (4.6 kg/m<sup>2</sup>)).

Actor/non-actor status was verified with self-report questionnaires assessing experience or training in acting or acting-related skills (voice & speech, movement, improvisation, etc) at 3 levels: taking classes, taking part in amateur productions, or taking part in professional productions (i.e. for which they received payment). Current or former Tisch students endorsed an average of 2.98 (S.D. 0.10) out of a maximum possible 3 levels; "Other" actors endorsed an

average of 2.50 (S.D. 0.50); and non-actor control participants endorsed an average of 0.22 (S.D. = 0.43). Comparisons of acting experience level between actors and controls (Tisch alone, or Tisch and other actors versus non-actor controls) were strongly different (two-sample t-tests,  $p$ 's < 5e-36).

There were no significant differences between non-actor controls and actors (either Tisch alone or Tisch and other actors) for weight or BMI. Actors were marginally more male than non-actors (including Tisch and other actors; 16/37 versus 6/29;  $X^2(1, N = 66) = 3.72, p = 0.054$ ), though significantly more Tisch actors specifically were male than were non-actor controls (14/26 versus 6/29;  $X^2(1, N = 55) = 6.51, p = 0.011$ ). Actors were not significantly older than non-actors ( $p = 0.19$ ), though Tisch actors were significantly older than non-actor controls ( $t(53) = 2.62, p = 0.012$ ; 95% CI [0.70, 5.27]). Actors were taller than non-actors (all actors vs. non-actor controls:  $t(64) = 3.05, p = 0.0034$ ; 95% CI [0.91 4.35]; Tisch actors vs. non-actor controls:  $t(53) = 4.8, p = 1.5e-5$ ; 95% CI [2.26, 5.54]). For analyses of state and trait anxiety, see Results: Questionnaires. For a summary of participant demographics (including sex, age, height, weight, BMI, state and trait anxiety, and acting experience level) organized by category (Tisch actors, other actors, and non-actor controls) see Table S1.

### *Heartbeat Measurement*

Heartbeats were measured using an electrocardiogram (ECG). We applied three electrodes to each participant (ground below left clavicle; shielded electrodes below right clavicle and below the ribcage in line with the ground electrode). Electrodes connected to a Biopac MP160 with an ECG100C module (1000Hz gain, norm mode, 35Hz LPN, high-pass filter at 1.0Hz), which itself

was connected to a dedicated physiology computer running AcqKnowledge software (version 4.1, Biopac Systems Inc, Goleta, CA). The software carried out automatic detection of R-waves (the peak of the ECG signal during ventricular depolarization) and subsequently triggered the experiment presentation software (PsychToolBox in MATLAB) on the stimulus computer to play tones at delays of 200ms or 500ms.

### *Experiment Structure*

Immediately after providing informed consent, participants were endowed with \$5 (see below), and then fully instructed in the heartbeat detection task, after which they completed a quiz assessing comprehension of task basics. Three electrodes were then applied for the ECG (see above, Heartbeat Measurement). Participants were instructed to relax and try to find their heartbeat for 5 silent minutes, during which time the experimenter left the room, and they were asked to refrain from checking their phones or using the computer. After 5 minutes, the experimenter re-entered the room and initiated the heartbeat detection task (see below). When the task finished, participants completed in-house questionnaires assessing their experience, as well as the State-Trait Anxiety Inventory (Spielberger, 1983). They were then paid and debriefed.

Participants were paid \$10/hr (for a total of \$15). To incentivize performance, participants were also told that a single trial would be randomly selected at the end of the study. If they had responded on that trial and were correct, they would receive \$5 in addition to their \$5 endowment (for a total bonus of \$10); if they had responded but were incorrect, they would simply keep the \$5 endowment; if they had not responded, they would lose the \$5 endowment.

*Heartbeat Detection Task*

The heartbeat detection task was a two-alternative forced choice task that has been extensively validated and used elsewhere (e.g. Critchley et al., 2004; Eichler & Katkin, 1994; Khalsa et al., 2008; Mulcahy et al., 2019; Sokol-Hessner et al., 2015; Wiens & Palmer, 2001). It is additionally considered a conservative measure of interoceptive sensitivity, thus our results are more likely to underestimate than overestimate true interoceptive effect sizes (see Supplementary Material; Brener & Ring, 2016). Each trial in the heartbeat detection task began with a 1s visual warning (“Attend” on the screen) that the trial was about to begin. Ten tones were then played at a low-to-moderate volume from speakers on either side of the computer monitor. All 10 tones on a given trial were played at the same delay from the participant’s R-wave (spike in the ECG). Delay values were chosen to correspond with the most-synchronous (R+ 200ms) and most-delayed (R+ 500ms) judgments as assessed in prior studies (Wiens et al., 2000). Good interoceptive sensitivity in this task corresponds to high accuracy in discriminating in-sync from delayed trials. After hearing the 10 tones, participants were asked whether the tones were in-sync or delayed from their heartbeat. The placement of “in-sync” and “delayed” options on the left or right side of the screen (and thus the mapping to the response buttons) varied randomly trial-by-trial. Participants had 2 seconds to enter their response. After every judgment of in-sync or delayed, participants were asked to indicate how confident they were in that judgment on a non-numerical analog scale from “low” (coded as 0) to “high” (coded as 1) by pressing one button to move the cursor left, and another button to move the cursor right. Cursor starting position was normally distributed trial-to-trial with a mean of 0.5 and a standard deviation of 0.04. Participants had 3 seconds to place the cursor where they wanted, after which it locked in place

for 0.5 seconds. Good metacognition corresponds to a positive relationship between confidence ratings and the probability of being correct in the interoceptive task on that trial. Following an ITI of 0.75s or 1.25s (equal probability), the next trial would begin.

The task began with 3 labeled in-sync trials interleaved with 3 labeled delayed trials (i.e. the screen said “This trial is in-sync” or “This trial is delayed”). Participants then completed 160 unlabeled and pseudorandomly intermixed trials (so that no more than 7 sequential trials were of the same type) separated across 4 blocks of 40 trials each. There was no feedback during the task. Taking into account our 66 participants (37 Actors and 29 Controls), we collected and analyzed a total of 10,560 interoception judgment trials.

### *Measures of Acting Ability and Acting Training*

Acting ability is inherently difficult to quantify, as it attempts to objectify performance, which is affected by the performer, the content being performed, the observer themselves, the observer’s subjective reaction to the performance, and interactions between these components. Training is similarly challenging to quantify, as it can vary significantly in intensity, duration, quality, and type.

As quantitative proxies for acting ability, we used three scores already being collected by the Tisch Graduate Acting Program to assess their students (two students did not release their scores; thus, for acting scores,  $N = 24$ ) that measured different aspects of body control/awareness and acting.

The “Mimicry” score assessed students’ body perception and awareness when not moving. The score was the average of two evaluators’ scores for a test in which students were given a photo of Rodin’s Thinker for 10 seconds, after which the photo was taken away, and students were asked to reproduce the sculpture’s positioning in detail from memory. This assessment was performed in the fall of their first year. Scores ranged from 0-100, with higher scores being better.

The second score, “Movement”, reflected students’ body control during movement in a time-pressured situation. The score reflected participants’ single best time out of 3 attempts at an obstacle course they had to move through as quickly as possible while manipulating and moving some objects, and not knocking down or disturbing other objects placed in their way. This assessment was performed in their most recent fall of graduate school. For consonance with the other scores (in which higher scores indicate better performance), we calculated the time difference for each subject between their personal best time and the overall slowest time observed across all participants, so that scores used in analysis ranged from 0-8.4s, with higher scores indicating faster performance.

The third score, “Audition,” reflected an integrated evaluation of students’ acting performance. The score was the average of two highly experienced auditors’ scores of participants’ initial audition for admission to the Tisch School of the Arts Graduate Acting Program. Scores ranged from 1 to 5 with lower scores being better. As our participant sample included only applicants who were subsequently admitted to Tisch Graduate Acting, scores for our participants ranged between 1 and 3. For consonance with the other scores (in which higher scores indicate better performance), we reversed the Audition scores for analysis so they ranged between zero and 2,

with higher scores being better. Note that the Audition score in particular was collected prior to training at Tisch Graduate Acting, and so may be best thought of as a measure of actors' initial ability before graduate school.

To quantify training, we calculated the number of months (rounded to the nearest month) between when participants matriculated at Tisch and when they participated in our study. Values ranged from 1-67 months (mean = 30.1 months, s.d. = 22.7 months). This measure of course does not account for training prior to enrollment in the Tisch Graduate Acting program.

### *Analysis Approach*

To maximize the statistical power to answer our main question of whether and how actors' interoceptive sensitivity and/or their metacognition about interoceptive performance differs from non-actor controls, we collapsed Tisch graduate actor participants and the "Other" actors into a single "Actor" group when possible and compared them to Controls.

Depending on the analytic question, we used a combination of t-tests (paired, two-sample, and one-sample), correlations, and 95% Bayesian credible intervals (with samples of parameter values resulting from model-fitting procedures, see below). Some of our central variables (including  $d'$ ) were distributed in ways that could bias correlation analyses. For simplicity, all reported correlations are therefore robust, nonparametric Spearman's Rho correlations. Analyses were carried out using MATLAB (version 9.4.0 (R2018a); Mathworks, Inc).

For accuracy in the heartbeat detection task, we calculated  $d'$ , a unitless Signal Detection Theory sensitivity measure capturing the degree to which participants could distinguish between in-sync and delayed trial types. Values of  $d'$  were calculated as the normalized hit rate minus the normalized false alarm rate.

Confidence ratings were analyzed to determine metacognitive awareness (metacognitive efficiency) in a bias-free framework that leveraged a signal detection theory approach directly analogous to the use of  $d'$  to quantify interoceptive sensitivity. Just as participants' judgments of "in-sync" or "delayed" were used to calculate sensitivity in interoception ( $d'$ ) on the basis of the correspondence between those judgments and the objective truth on that trial (whether it was actually "in-sync" or "delayed"), confidence ratings were used to calculate metacognitive sensitivity about interoceptive accuracy (meta- $d'$ ) on the basis of the correspondence between confidence ratings and the objective truth on that trial (whether participants were correct or incorrect in their interoceptive judgment).

Analysis was carried out in JAGS (Just Another Gibbs Sampler; Plummer, 2003), a Markov-chain Monte Carlo (MCMC) hierarchical Bayesian estimation tool, as implemented in matJAGS. Sampling-based hierarchical Bayesian estimation techniques excel at fitting complex non-linear models in studies like ours, especially given the number of parameters in our model (separate metacognitive parameters for each of our participants plus the group-level parameters that govern the distribution of the individuals' values; see below), and the hierarchical dependencies between them (see Equations 1 and 2). Simpler models that are linear, do not leverage dependencies between data, and/or have fewer parameters can be fit effectively with classic

maximum likelihood estimation techniques that seek the single best-fitting parameter estimate(s). However, more complex models like ours are most efficiently and effectively fit using MCMC techniques.

The modeling approach was based upon the receiver-operating-characteristic (ROC) framework which required that confidence ratings were quantized, in this case using quartiles to replace continuous ratings with discrete ratings from 1-4 (see Fleming & Lau, 2014 for a detailed explanation of the mathematics of this approach, and a discussion of the issues of bias in measuring metacognition). Briefly, this modeling approach does not model metacognitive sensitivity ( $meta-d'$ ) directly because metacognitive sensitivity is expected to be constrained by performance (Maniscalco & Lau, 2012). Thus, metacognition was modeled as the *ratio* between  $meta-d'$  and  $d'$  ( $meta-d'/d'$ , referred to as “M-ratio” or metacognitive efficiency). Higher values of M-ratio (closer to 1) indicate metacognitive efficiency closer to optimal levels (i.e. 1).

In the standard hierarchical meta- $d'$  (HMeta- $d'$ ) model (Fleming, 2017), participants’ individual M-ratio parameter values ( $\mu_i$  where the  $i$  subscript represents different participants) are distributed around a group-level normal distribution, as in Equation 1.

$$\mu_i = normal(\mu_M, \mu_{SD}) \quad \text{Equation 1}$$

The group-level normal distribution itself is parameterized with a mean M-ratio value ( $\mu_M$ ) and standard deviation ( $\mu_{SD}$ ). This hierarchical approach enables all participants’ data to be modeled and fit simultaneously, estimating individuals’ parameter values in the context of our knowledge of the entire group and all data points, thereby leveraging that knowledge to maximize signal and minimize the influence of noise.

We modified the classic hierarchical meta-d' model to quantify a potential improvement in Actors' metacognitive efficiency with a parameter (the potential "Actors' Bonus") added to all Actors' individual M-ratio values in a joint model of both Actors and Controls, as in Equation 2.

$$\mu_i = \text{normal}(\mu_M, \mu_{SD}) + \beta * A_i \quad \text{Equation 2}$$

In this equation,  $\beta$  represents the Actor's Bonus parameter value, multiplied by a dummy variable  $A_i$  which has value 1 when participant  $i$  is an Actor, and 0 when they are a Control participant. In this framework,  $\beta$  essentially functions as a difference score, capturing the mean difference between the metacognitive efficiency of Actors and Controls in a single parameter estimate. This approach allows the specification of a prior on that difference of interest and enables the direct examination of the sampled values of that term to ascertain whether and how Actors' and Controls' metacognitive efficiency differs. In short, this approach leverages our knowledge that some participants are Actors and others are Controls by directly incorporating that potential difference into the structure of the model, while also acknowledging and fitting individual differences.

Our main analysis thus examines the distribution of sampled values of the Actor's Bonus parameter, and in particular, whether that distribution appears to be consistently positive.

For each of 4 MCMC chains, we used burn-in periods of 1,000 samples. These samples were discarded to reduce the influence of both our selections of hyperprior distributions for the group-level parameters and the random starting points for the sampling chains. After the burn-in period in each chain, 10,000 samples were collected, for a total of 40,000 samples across the four

chains. Hyperpriors for the parameters were selected to be uninformative - see Supplementary Material for detailed discussion of hyperpriors (note too that the burn-in period is intended to minimize their influence). Given uninformative priors and the discarded burn-in period, the final set of 40,000 samples are best thought of as reflecting the likelihood of the data given our estimates and model. Estimates of convergence across chains were good, with R-hat convergence values for the three group-level parameters of  $\mu_M$ ,  $\mu_{SD}$ , and  $\beta$  all  $< 1.005$  (below standard cutoffs; Vehtari et al., 2021). Visual inspection of the overlap between the fitted and observed Type 2 (metacognitive) ROC curves for all participants (as well as Actors alone and Controls alone) indicated an excellent level of overlap (see Figure S1), suggesting that the model accurately fit participants' metacognitive judgments. 95% credible intervals were examined for the values of each parameter of interest using the final 40,000 samples.

#### *Data and Model Files*

De-identified data and model files for the actor's bonus hierarchical *meta-d'* model are available via the Open Science Framework at <https://osf.io/84cpk/>.

### **III. Results**

#### *Heartbeat Detection Task: Interoception*

Overall, participants' average performance in the heartbeat detection task was significantly better than chance (mean  $p(\text{correct}) = 0.62$ ,  $SE = 0.12$ ; one sample t-test against 0.5,  $t(65) = 7.7$ ,  $p = 1.1 \times 10^{-10}$ ). While Actors had numerically higher performance (mean  $p(\text{correct}) = 0.64$ ,  $SE = 0.13$ ) than Controls (mean  $p(\text{correct}) = 0.59$ ,  $SE = 0.12$ ), the difference was not significant ( $t(64)$

= 1.4,  $p = 0.17$ ; 95% CI [-0.018, 0.10]), contrary to hypotheses that Actors would have significantly better interoceptive sensitivity compared to non-actor Controls.

All subsequent analyses (which leverage signal detection theory) use the unitless sensitivity index of  $d'$  (d-prime), which is highly correlated with  $p(\text{correct})$  (Spearman's Rho,  $r(64) = 0.998$ ,  $p = 6.1 \times 10^{-77}$ ).  $D$ -prime was uncorrelated with height, weight, BMI, age, "state" anxiety (STAI-S), or "trait" anxiety (STAI-T) (Spearman's correlations; all  $p$ 's  $> 0.1$ ).

#### *Heartbeat Detection Task: Metacognition*

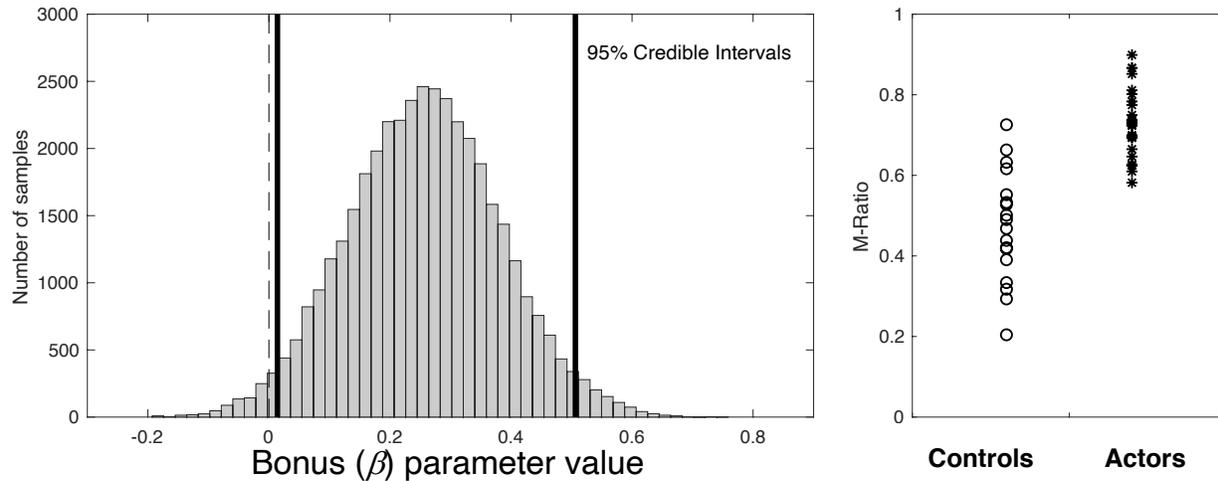
Actors and Controls did not significantly differ in their average absolute confidence levels (often referred to as metacognitive bias) during the interoception task (mean [standard error of the mean] confidence ratings; Actors = 0.60 [0.03], Controls = 0.56 [0.03]; two-sample  $t$ -test,  $p = 0.27$ ). However, while average levels of confidence may indicate that people use confidence scales differently or are more or less confident in general, averages do not capture the hallmark of metacognitive sensitivity: trial-by-trial changes in confidence that track trial-by-trial changes in accuracy. In individuals with high levels of metacognitive ability, trials with high confidence are much more likely to be correct than trials on which their confidence is low. This contrasts with someone with low or no metacognitive ability who will just as likely to be correct when their confidence is high or low. Good metacognitive performance can thus be thought of as resulting from a close *correspondence* between accuracy and confidence on a per-trial basis. Because metacognitive sensitivity is normatively expected to scale with performance itself (i.e. meta- $d'$  should equal  $d'$  under signal detection theory; Maniscalco & Lau, 2012), we can

summarize each individual's performance-corrected metacognitive "efficiency" as the ratio  $\text{meta-}d'/d'$ , known as "M-ratio".

To quantitatively estimate metacognitive efficiency we used a hierarchical Bayesian model, fit using Markov chain Monte Carlo (MCMC) sampling methods (see Methods) to estimate group-level M-ratios. Because metacognition is expected to be constrained by performance (i.e. with near chance-level performance, there is limited opportunity to demonstrate successful metacognition), we removed participants who performed at or near chance levels on the interoception task from this analysis (see Supplementary Materials for results of a model using all participants). Using a  $d'$  threshold value of 0.25 (corresponding to roughly 54% correct), 21 participants were removed (6 Tisch Graduate Actors, 4 Other Actors, and 11 Controls), leaving 45 participants for the analysis of metacognition (20 Tisch Graduate Actors, 7 Other Actors, and 18 Controls). There were still no significant differences between Actors and Controls in performance in the heartbeat detection task after these exclusions (two-sample t-test on  $p(\text{correct})$ ,  $t(43) = 0.73$ ,  $p = 0.47$ ; on  $d'$ ,  $t(43) = 0.92$ ,  $p = 0.36$ ).

To assess whether Actors had better metacognitive sensitivity about interoceptive accuracy than Controls, we examined the MCMC samples of the "Bonus" ( $\beta$ ) parameter added to all Actors' M-Ratios (see Figure 1). Consistent with hypotheses, the mean value of  $\beta$  was 0.257, indicating a net positive effect such that Actors had a higher level of metacognitive efficiency than non-actor Controls. The 95% credible interval of the samples of the Bonus parameter (0.014, 0.506) was also reliably positive, indicating that given the observed data, samples of the Actor's Bonus parameter values had a 95% probability of falling within this positive range. Fully 98.1% of

sampled values were also above zero. This suggests that Actors had consistently better metacognitive awareness of interoceptive accuracy compared to non-actor Controls.



[Figure 1: Histogram of samples of the Actors' Bonus ( $\beta$ ) parameter and plot of individual participant M-Ratio values. The histogram reflects 40,000 MCMC samples of the metacognitive efficiency bonus term added to Actors' M-Ratio values. Thick black lines indicate the 95% CIs of the samples. On the right side, final M-Ratio values (including baseline values and the Actors' Bonus term for Actors) are plotted separately for Controls (o) and Actors (\*). An M-Ratio of 1 indicates optimal metacognitive performance with respect to interoceptive performance, while an M-Ratio of 0 indicates no metacognitive insight.]

Exploratory analyses examined whether improved metacognitive efficiency for interoceptive accuracy was related to acting ability by correlating each actor's M-Ratio value (the sum of the Actor's Bonus and their individual M-Ratio) with the three acting ability scores available for Tisch Graduate Acting students (Mimicry, Movement, and Audition) using robust nonparametric Spearman's Rho correlations. As not all Actors were from Tisch, and not all Tisch students released their scores, this analysis was only able to leverage 19 participants' data – thus, these should be considered preliminary findings, and interpreted with caution. We found that Actors'

M-Ratios were correlated with Mimicry and Movement scores (with Mimicry,  $r(17) = 0.48$ ,  $p = 0.04$ ; with Movement,  $r(17) = 0.49$ ,  $p = 0.03$ ; note that Mimicry and Movement scores were themselves correlated, see below). There was no correlation with Audition scores ( $r(17) = -0.08$ ,  $p = 0.73$ ).

To examine whether the amount of acting training actors had received correlated with improved metacognitive efficiency for interoceptive accuracy, we also correlated metacognitive ability with an estimate of time spent training (quantified as the number of months since beginning the Tisch Graduate Acting program), but found no significant relationship ( $r(17) = 0.05$ ,  $p = 0.83$ ).

When examining metacognitive awareness of interoceptive accuracy overall (across Actors and Controls) using robust nonparametric Spearman's correlations, metacognitive efficiency was significantly correlated with height ( $r(43) = 0.56$ ,  $p = 5.6 \times 10^{-5}$ ; taller individuals had better metacognition, though note that Actors were significantly taller than Controls, both within the metacognition-analysis group [ $N=45$ ;  $t(43) = 4.7$ ,  $p = 2.4 \times 10^{-5}$ ; 95% CI: [2.41, 5.98]] and overall [ $N = 66$ ;  $t(64) = 3.0$ ,  $p = 0.003$ ; 95% CI: [0.91, 4.35]]) and marginally correlated with age ( $r(43) = 0.29$ ,  $p = 0.051$ ; age was not significantly different between Actors and Controls with either  $N = 45$  or  $N = 66$ ,  $p$ 's > 0.19). No correlation was observed between metacognitive efficiency estimates and weight, BMI, STAI-S, or STAI-T (all  $p$ 's > 0.25). Finally, metacognitive efficiency was also uncorrelated with interoceptive sensitivity (with  $d'$ ;  $r(43) = -0.04$ ,  $p = 0.80$ ), suggesting that metacognition about interoceptive accuracy relies on distinct mechanisms to those supporting interoceptive sensitivity itself.

Finally, all participants self-reported their estimated overall probability of being correct on the heartbeat detection task at the conclusion of the study as a measure of “global” metacognition about performance (from 50-100%; this was of course highly negatively correlated with their rating of the difficulty of the task on a scale from 1 [very easy] to 7 [very difficult]; Spearman’s Rho,  $r(64) = -0.52$ ,  $p = 6.5 \times 10^{-6}$ ). Although participants slightly overestimated their overall probability of being correct on average ( $M_{\text{estimated}} = 0.66$ ;  $M_{\text{actual}} = 0.62$ ; paired t-test,  $t(65) = 2.70$ ,  $p = 0.009$ , 95% CI: [0.012 0.078]), their actual and estimated  $p(\text{correct})$  were significantly correlated (Spearman’s Rho;  $r(64) = 0.30$ ,  $p = 0.01$ ), suggesting a surprisingly high overall degree of post-hoc global metacognitive awareness of their interoceptive performance, despite a lack of feedback during the task. There was no significant difference between Actors and Controls in the  $p(\text{correct})$  judgment error (because this quantity was obviously not normally distributed, we used the nonparametric Wilcoxon rank sum test; signed difference between actual and estimated  $p(\text{correct})$  for Actors vs. Controls,  $p = 0.72$ ). While relationships between actual performance and later global estimation of performance have been established in other perceptual domains (Rouault et al., 2019; Rouault & Fleming, 2020), this is the first evidence of which we are aware for the same phenomenon in the interoceptive domain.

### *Questionnaires*

Post-task ratings of the difficulty of the heartbeat detection task on a scale from 1 (very easy) to 7 (very difficult), revealed that Actors found the task easier ( $M = 4.90$ ,  $SE = 1.66$ ) than non-actor Controls ( $M = 5.93$ ,  $SE = 0.84$ ; two-sample t-test,  $t(64) = 3.1$ ,  $p = 0.003$ ; 95% CI [0.36, 1.72]). There were no significant differences in the difficulty of making confidence ratings between Actors ( $M = 3.8$ ,  $SE = 1.6$ ) and Controls ( $M = 4.2$ ,  $SE = 1.6$ ;  $p = 0.36$ ).

While Actors and Controls were only marginally different on “state” anxiety (STAI-S; Actors  $M = 37.5$ ,  $SE = 9.7$ , Controls  $M = 33.5$ ,  $SE = 7.7$ ,  $t(64) = 1.8$ ,  $p = 0.07$ ; 95% CI: [-0.40, 8.39]), Actors were significantly more “trait” anxious than Controls (STAI-T; Actors  $M = 45.3$ ,  $SE = 10.9$ , Controls  $M = 39.7$ ,  $SE = 10.7$ ,  $t(64) = 2.09$ ,  $p = 0.04$ ; 95% CI: [0.25, 11.0]).

Scores of acting ability were available for almost all current or former Tisch Graduate Acting students ( $N = 24$ ; see Methods). Two of the measures were positively correlated with each other (Mimicry and Movement;  $r(22) = 0.46$ ,  $p = 0.02$ ), possibly suggesting a shared basis, as both tests emphasize bodily control. Neither Mimicry nor Movement were correlated with Audition scores (both  $p$ 's  $> 0.4$ ). As these scores themselves are novel, we performed exploratory analyses to examine whether they were related to attributes of the Actors. Males scored marginally better on Movement ( $t(22) = 1.8$ ,  $p = 0.08$ ), while there were no differences between sexes for the other assessments (Mimicry  $p = 0.2$ ; Audition  $p = 0.41$ ). Using robust nonparametric Spearman's Rho correlations, no relationships were observed between any measures of acting ability and height, weight, body mass index (BMI), age, STAI-S, or STAI-T (all  $p$ 's  $> 0.11$ ).

#### **IV. Conclusions**

Here we robustly quantify and dissociate the perception of cardiac interoceptive signals (interoceptive sensitivity) from knowledge of that perception (metacognition about interoceptive accuracy), finding that these two abilities are independent in this domain. While null results should be interpreted with caution, our findings are consistent with the conclusion that accuracy and insight in interoception are separable constructs. We additionally find that while Actors do

not have significantly greater interoceptive sensitivity than non-actors, they do appear to have consistently greater metacognitive sensitivity to interoceptive accuracy. Moreover, this increased metacognitive insight may be related to some proxy measures of acting ability, though not to an estimate of the amount of acting training they have received. Finally, we identified a consistently high level of post-hoc global metacognitive ability that, in contrast to metacognitive efficiency during the task, did not differ between Actors and Controls.

Identifying who differs in interoceptive accuracy and/or metacognitive sensitivity to interoceptive accuracy, and in what ways is the first step toward understanding how and why these differences arise and their consequences. We recruited actors to address this question because of the centrality of accurate self-perception and awareness to their profession generally, and especially within the domain of physiology and emotion (Gosselin et al., 1995; Stanislavsky, 1952; Strasberg, 1987). Because of the established connection between interoception and emotional experience (Barrett et al., 2004; Critchley et al., 2004; Khalsa et al., 2018; Wiens et al., 2000), empathy (Ernst et al., 2013; Heydrich et al., 2021; Terasawa et al., 2014), and emotion regulation (Kever et al., 2015), we think it likely that the enhanced metacognitive awareness in actors identified here will facilitate actors' emotion-related abilities, as the awareness of the perception of internal physiological signals would likely be relevant to the ability to effectively create, express, and convey emotion as an actor, though this is of course an empirical question. There is also the potential for these factors to be additionally related to other important emotion constructs like emotional awareness or intelligence (e.g. Ashkanasy & Dasborough, 2003; Lane et al., 1990; Salovey & Grewal, 2005), emotion expression (e.g. Elfenbein et al., 2002;

Matsumoto et al., 2008), or affective synchrony (e.g. Wood et al., 2021) which could be examined by future studies.

The metrics of acting ability used here focused on bodily control and general performance abilities, and while these would be relevant for the ability to convey and express emotion, they are not themselves measures of emotion in acting. Nevertheless, the relationships we did observe suggest that other metrics of acting-specific abilities and/or outcomes, like audience reactions, intensity or consistency of expressed emotion, perceivers' elicited emotions, and more might also be related to metacognitive sensitivity to interoceptive accuracy. We hope that future research will explore these rich and engaging possibilities.

That actors have better metacognitive awareness of interoceptive ability raises the question of how they came to be that way. The two main possibilities are that acting training and practice have honed that ability, or that people with better metacognitive insight to interoceptive accuracy disproportionately become and succeed as actors. While interoceptive sensitivity is often considered a stable or trait-like ability (e.g. Khalsa et al., 2008), there is some evidence that metacognitive ability can be improved with training, at least in non-emotional domains (Carpenter et al., 2019). Although we found no relationship between the time Tisch actors had been in graduate school and their metacognitive ability, it is possible that we did not sample a wide enough range of training levels with our measure, given the consistently high quality of Tisch Graduate Acting students (Abramovitch, 2021). In other words, our measure of training (time since matriculating at Tisch) may not be an accurate quantification of acting training. Without longitudinal data, a wider range of training levels, different types of acting training,

and/or better quantification of training duration, we cannot distinguish between the possibilities that training improves metacognition about interoceptive accuracy or that there is a selection effect in which only individuals with high metacognitive ability become actors. Nevertheless, these data establish for the first time that there is a quantitative, objective difference in the first place, and suggest it is not related to the amount of time actors have been training, though future research will have to test this empirically and/or with different samples.

While actors had consistently better metacognitive awareness of interoceptive accuracy, it is unclear whether this is part of a domain-general enhanced metacognitive ability (in which case actors should have better metacognitive awareness of their mnemonic or perceptual abilities, for example), or whether the improvement is domain-specific, limited to interoception. The latter seems most likely, with evidence that metacognitive ability appears to develop independently across domains in childhood (Vo et al., 2014), that domain-specific deficits in metacognition have been observed following brain damage (Fleming et al., 2014), and that separate psychiatric symptom patterns are linked to unique changes in metacognitive ability (Rouault et al., 2018). Future research will have to address these possibilities. Because the analysis of metacognitive sensitivity to interoceptive accuracy was also limited to those participants who demonstrated some positive level of interoceptive accuracy in the current task, we must also leave questions regarding those participants categorized by this task as relatively poor interoceptors (for any reason) to future research.

Until very recently, it has proven difficult to gather enough interoceptive performance data to robustly estimate the link between accuracy and confidence, as unlike in other perceptual

modalities, it has often been unfeasible to collect interoceptive data over hundreds of trials. By using a relatively high-powered design (160 trials per subject; more than double that of any other interoception study of which we are aware) and hierarchical Bayesian modeling, we robustly separated interoceptive sensitivity from metacognitive sensitivity to interoceptive accuracy (see Supplementary Material). We are aware of only one other study that has taken a similar analytic approach (Harrison et al., 2021), using hierarchical modeling to estimate metacognitive sensitivity to interoception in the domain of respiration interoception, and comparing individuals with and without asthma (finding no differences between the groups in any interoceptive or metacognitive measures). Other prior research has examined the degree of coherence between different affective measurements (e.g. between cardiac activity and ratings of positive/negative affect on a likert scale; Sze et al., 2010) or used simpler metrics of metacognitive sensitivity (e.g. Chua & Bliss-Moreau, 2016; Forkmann et al., 2016; Garfinkel et al., 2014; Mulcahy et al., 2019), but only recently has it become possible to leverage the toolbox of hierarchical Bayesian modeling and signal detection theory to effectively separate task accuracy from metacognitive efficiency in a range of domains.

As we identify complex emotion-cognition interactions in many domains (Dolan, 2002), it becomes increasingly important to understand the factors that contribute to our emotional experience, including how we perceive and know about our own bodies, as these constructs will contribute not only to our emotions, but to how we interact with them. Our powerful approach showed not only that accurately perceiving our own body and knowing that we can do that are two different things, but that there is the potential for special populations like actors to have objectively better insight into the accuracy of their perceived bodily experience.

**Author Contributions:**

MWD, SI, and EAP initiated the collaboration. PSH developed the study concept in collaboration with SI and MWD. All authors contributed to the design of the study. PSH collected the data. PSH and SMF analyzed the data under the supervision of EAP. PSH and EAP drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

**Acknowledgments:**

The authors thank Jackie Reitzes for assistance with data collection.

## References

- Abramovitch, S. (2021, June 19). The World's 25 Best Drama Schools, Ranked. *The Hollywood Reporter*. <https://www.hollywoodreporter.com/lists/25-best-drama-schools-ranked/>
- Ashkanasy, N. M., & Dasborough, M. T. (2003). Emotional Awareness and Emotional Intelligence in Leadership Teaching. *Journal of Education for Business*, 79(1), 18–22. <https://doi.org/10.1080/08832320309599082>
- Ateş Çöl, I., Sönmez, M. B., & Vardar, M. E. (2016). Evaluation of Interoceptive Awareness in Alcohol-Addicted Patients. *Noro Psikiyatri Arsivi*, 53(1), 17–22. <https://doi.org/10.5152/npa.2015.9898>
- Barrett, L. F., Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive Sensitivity and Self-Reports of Emotional Experience. *Journal of Personality and Social Psychology*, 87(5), 684–697. <https://doi.org/10.1037/0022-3514.87.5.684>
- Bonaz, B., Lane, R. D., Oshinsky, M. L., Kenny, P. J., Sinha, R., Mayer, E. A., & Critchley, H. D. (2021). Diseases, Disorders, and Comorbidities of Interoception. *Trends in Neurosciences*, 44(1), 39–51. <https://doi.org/10.1016/j.tins.2020.09.009>
- Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160015–20160019. <https://doi.org/10.1098/rstb.2016.0015>
- Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, 148(1), 51–64. <https://doi.org/10.1037/xge0000505>

- Chua, E. F., & Bliss-Moreau, E. (2016). Knowing your heart and your mind: The relationships between metamemory and interoception. *Consciousness and Cognition, 45*(C), 146–158. <https://doi.org/10.1016/j.concog.2016.08.015>
- Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7*(2), 189–195.
- Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science, 298*(5596), 1191–1194. <https://doi.org/10.1126/science.1076358>
- Dumas, D., Doherty, M., & Organisciak, P. (2020). The psychology of professional and student actors: Creativity, personality, and motivation. *PLOS ONE, 15*(10), e0240728. <https://doi.org/10.1371/journal.pone.0240728>
- Eichler, S., & Katkin, E. S. (1994). The relationship between cardiovascular reactivity and heartbeat detection. *Psychophysiology, 31*, 229–234.
- Elfenbein, H. A., Marsh, A. A., & Ambady, N. (2002). Emotional intelligence and the recognition of emotion from facial expressions. In *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 37–59). The Guilford Press.
- Ernst, J., Northoff, G., Böker, H., Seifritz, E., & Grimm, S. (2013). Interoceptive awareness enhances neural activity during empathy. *Human Brain Mapping, 34*(7), 1615–1624. <https://doi.org/10.1002/hbm.22014>
- Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness, 2017*(1), 377–14. <https://doi.org/10.1093/nc/nix007>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 16657–30. <https://doi.org/10.3389/fnhum.2014.00443>

- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811–2822. <https://doi.org/10.1093/brain/awu221>
- Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., & Schulz, A. (2016). Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology*, *109*, 71–80. <https://doi.org/10.1016/j.ijpsycho.2016.09.019>
- Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2014). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, *104*, 1–10. <https://doi.org/10.1016/j.biopsycho.2014.11.004>
- Garfinkel, S. N., Tiley, C., O’Keeffe, S., Harrison, N. A., Seth, A. K., & Critchley, H. D. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychology*, *114*, 117–126. <https://doi.org/10.1016/j.biopsycho.2015.12.003>
- Goldstein, T. R., & Bloom, P. (2011). The mind on stage: Why cognitive scientists should study acting. *Trends in Cognitive Sciences*, *15*(4), 141–142. <https://doi.org/10.1016/j.tics.2011.02.003>
- Goldstein, T. R., & Lerner, M. D. (2018). Dramatic pretend play games uniquely improve emotional control in young children. *Developmental Science*, *21*(4), e12603. <https://doi.org/10.1111/desc.12603>
- Gosselin, P., Kirouac, G., & Doré, F. Y. (1995). Components and Recognition of Facial Expression in the Communication of Emotion by Actors. *Journal of Personality and Social Psychology*, *68*(1), 83–96.

Gray, M. A., & Critchley, H. D. (2007). Interoceptive Basis to Craving. *Neuron*, *54*(2), 183–186.

<https://doi.org/10.1016/j.neuron.2007.03.024>

Harrison, O. K., Garfinkel, S. N., Marlow, L., Finnegan, S. L., Marino, S., Köchli, L., Allen, M., Finnemann, J., Keur-Huizinga, L., Harrison, S. J., Stephan, K. E., Pattinson, K. T. S., & Fleming, S. M. (2021). The Filter Detection Task for measurement of breathing-related interoception and metacognition. *Biological Psychology*, *165*, 108185.

<https://doi.org/10.1016/j.biopsycho.2021.108185>

Heydrich, L., Walker, F., Blättler, L., Herbelin, B., Blanke, O., & Aspell, J. E. (2021).

Interoception and Empathy Impact Perspective Taking. *Frontiers in Psychology*, *11*, 4016. <https://doi.org/10.3389/fpsyg.2020.599429>

Jackson, D., & Muir, K. (2019). Individual differences in acted emotions: An exploratory study into the role of interoceptive awareness in feelings prompted by emotion eliciting acting techniques. In *PsyArXiv*. <https://doi.org/10.31234/osf.io/shf72>

Kandasamy, N., Garfinkel, S. N., Page, L., Ben Hardy, Critchley, H. D., Gurnell, M., & Coates, J. M. (2016). Interoceptive Ability Predicts Survival on a London Trading Floor. *Scientific Reports*, *6*(1), 1–7. <https://doi.org/10.1038/srep32986>

Keiver, A., Pollatos, O., Vermeulen, N., & Grynberg, D. (2015). Interoceptive sensitivity facilitates both antecedent- and response-focused emotion regulation strategies. *Personality and Individual Differences*, *87*, 20–23.

<https://doi.org/10.1016/j.paid.2015.07.014>

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P.,

- Simmons, W. K., Stein, M. B., ... Zucker, N. (2018). Interoception and Mental Health: A Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(6), 1–13. <https://doi.org/10.1016/j.bpsc.2017.12.004>
- Khalsa, S. S., Rudrauf, D., Damasio, A. R., Davidson, R. J., Lutz, A., & Tranel, D. (2008). Interoceptive awareness in experienced meditators. *Psychophysiology*, 45(4), 671–677. <https://doi.org/10.1111/j.1469-8986.2008.00666.x>
- Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A Cognitive-Developmental Measure of Emotion. *Journal of Personality Assessment*, 55(1–2), 124–134. <https://doi.org/10.1080/00223891.1990.9674052>
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. <https://doi.org/10.1016/j.concog.2011.09.021>
- Matsumoto, D., Keltner, D., Shiota, M. N., O’Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In *Handbook of emotions, 3rd ed* (pp. 211–234). The Guilford Press.
- Mulcahy, J. S., Davies, M., Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2019). Interoceptive awareness mitigates deficits in emotional prosody recognition in Autism. *Biological Psychology*, 146, 107711. <https://doi.org/10.1016/j.biopsycho.2019.05.011>
- Noice, H., & Noice, T. (2006). What Studies of Actors and Acting Can Tell Us About Memory and Cognitive Functioning. *Current Directions in Psychological Science*, 15(1), 14–18.
- Plummer, M. (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. 10.

- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1141.  
<https://doi.org/10.1038/s41467-019-09075-3>
- Rouault, M., & Fleming, S. M. (2020). Formation of global self-beliefs in the human brain. *Proceedings of the National Academy of Sciences*, *117*(44), 27268–27276.  
<https://doi.org/10.1073/pnas.2003094117>
- Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *BPS*, *84*(6), 1–9. <https://doi.org/10.1016/j.biopsycho.2017.12.017>
- Salovey, P., & Grewal, D. (2005). The Science of Emotional Intelligence. *Current Directions in Psychological Science*, *14*(6), 281–285. <https://doi.org/10.1111/j.0963-7214.2005.00381.x>
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.
- Sokol-Hessner, P., Hartley, C. A., Hamilton, J. R., & Phelps, E. A. (2015). Interoceptive ability predicts aversion to losses. *Cognition & Emotion*, *29*(4), 695–701.  
<https://doi.org/10.1080/02699931.2014.925426>
- Sokol-Hessner, P., Wing-Davey, M., Illingworth, S., Fleming, S.M., Phelps, E.A. (2021, November 8). The Actor's Insight: Actors have comparable interoception but better metacognition than non-actors. Retrieved from [osf.io/84cpk](https://osf.io/84cpk)
- Spielberger, C. D. (1983). State-trait anxiety inventory for adults. <https://doi.org/10.1037/t06496-000>
- Stanislavsky, K. (1952). *My Life in Art*. Taylor & Francis.

- Stewart, J. L., Juavinett, A. L., May, A. C., Davenport, P. W., & Paulus, M. P. (2015). Do you feel alright? Attenuated neural processing of aversive interoceptive stimuli in current stimulant users. *Psychophysiology*, *52*(2), 249–262. <https://doi.org/10.1111/psyp.12303>
- Stewart, J. L., May, A. C., Tapert, S. F., & Paulus, M. P. (2015). Hyperactivation to pleasant interoceptive stimuli characterizes the transition to stimulant addiction. *Drug and Alcohol Dependence*, *154*, 264–270. <https://doi.org/10.1016/j.drugalcdep.2015.07.009>
- Strasberg, L. (1987). *A dream of passion: The development of the method*. Boston : Little, Brown. [http://archive.org/details/dreamofpassionde0000stra\\_p6u7](http://archive.org/details/dreamofpassionde0000stra_p6u7)
- Sze, J. A., Gyurak, A., Yuan, J. W., & Levenson, R. W. (2010). Coherence between emotional experience and physiology: Does body awareness training have an impact? *Emotion*, *10*(6), 803–814. <https://doi.org/10.1037/a0020146>
- Terasawa, Y., Moriguchi, Y., Tochizawa, S., & Umeda, S. (2014). Interoceptive sensitivity predicts sensitivity to the emotions of others. *Cognition and Emotion*, *28*(8), 1435–1448. <https://doi.org/10.1080/02699931.2014.888988>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $R^{\hat{}}$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young Children Bet on Their Numerical Skills: Metacognition in the Numerical Domain. *Psychological Science*. <https://doi.org/10.1177/0956797614538458>

Wiens, S., Mezzacappa, E. S., & Katkin, E. S. (2000). Heartbeat detection and the experience of emotions. *Cognition & Emotion, 14*(3), 417–427.

<https://doi.org/10.1080/026999300378905>

Wiens, S., & Palmer, S. N. (2001). Quadratic trend analysis and heartbeat detection. *Biological Psychology, 58*, 159–175.

Wood, A., Lipson, J., Zhao, O., & Niedenthal, P. (2021). Forms and Functions of Affective Synchrony. In M. D. Robinson & L. E. Thomas (Eds.), *Handbook of Embodied Psychology: Thinking, Feeling, and Acting* (pp. 381–402). Springer International Publishing. [https://doi.org/10.1007/978-3-030-78471-3\\_17](https://doi.org/10.1007/978-3-030-78471-3_17)

**Supplementary Material** for “The Actor’s Insight: Actors have comparable interoception but better metacognition than non-actors”

Sokol-Hessner, P., Wing-Davey, M., Illingworth, S., Fleming, S.M., Phelps, E.A.

### Streaky Performance

Many participants reported post-hoc that their performance felt “streaky” in that their probability of being correct waxed and waned over the course of the study (e.g. that they would perform better for a few trials, then worse for a few trials). We analyzed this using the expected and observed numbers of repetitions in performance (e.g. the number of times they were correct on back-to-back trials, or incorrect on back-to-back trials). If participants were “streaky”, we would expect unusually high numbers of repetitions in performance. We used participants’ overall probability of being correct to construct individualized estimates of the expected probability of repetitions in performance, assuming no “streakiness” (i.e.  $p(\text{correct}) \times p(\text{correct}) + p(\text{incorrect}) \times p(\text{incorrect})$ ). We compared that estimated probability of repetition to the actual observed likelihood of a repetition in performance ( $p(\text{correct}_t | \text{correct}_{t-1}) + p(\text{incorrect}_t | \text{incorrect}_{t-1})$ ), in which  $t$  = current trial and  $t-1$  = previous trial), using paired t-tests to identify any consistent deviation between observed and expected probability of repetition.

We found no evidence of “streaky” performance. Using a paired-samples t-test across all participants to compare each participant’s observed likelihood of repetitions in performance with the expected probability of those repetitions given that participant’s overall task performance, we identified no consistent deviation between observed and expected probability of performance repetition ( $p = 0.67$ ).

### Measures of Response Bias

In the signal detection theory,  $d'$  represents “sensitivity”, or the degree to which observers distinguish between two different stimuli in their responses. In this framework, observers’ distinct behavioral responses result from subjectively distinct internal representations elicited by these objectively different stimuli. This sensitivity is distinguished from the criterion, or the level of subjective response above which individuals choose one category, and below which they choose another. A given criterion, in combination with a given sensitivity, results in a particular observed pattern of hits and false alarms. Changing the criterion changes the balance between hits and false alarms, but critically does not alter sensitivity. Put another way, sensitivity and criterion are theoretically independent (except in truly extreme cases of response bias in which responses approach 100% of one category and 0% of the other). The signal detection theory framework allows the assessment of sensitivity ( $d'$ ) separately from the criterion (a measure of response bias), which is a major benefit of using this framework.

In the current study, this applies to both our measure of interoceptive performance ( $d'$ , the sensitivity with which participants could distinguish between in-sync and delayed trials in the heartbeat detection task as measured by their in-sync/delayed responses), and metacognitive awareness of interoceptive performance (meta- $d'$ , the sensitivity with which participants could internally distinguish between trials in which they were correct and trials in which they were guessing, as measured by confidence ratings on those trial types). This is one of the major benefits of the metacognitive framework used here – that it is (in theory) *bias-free*.

Despite our focus on sensitivity, we were also able to measure response bias. In the interoception task, if “delayed” judgments were coded as 0, and “in sync” judgments were coded as 1, the average judgment was 0.58 across all participants (0.5 would indicate no net bias; one-sample t-test against 0.5:  $t(65) = 7.5$ ,  $p = 2.5e-10$ ). This was nearly identical across Actors ( $M = 0.59$ ) and Controls ( $M = 0.58$ ; two-sample t-test,  $p = 0.70$ ), indicating that there was no consistent difference in response bias across our main groups of interest. As discussed in the main text, average confidence ratings (metacognitive bias) also did not significantly differ between Actors and Controls.

That there was no difference between groups in the net bias in either interoception judgments or confidence ratings highlights the power of our trial-wise analysis approach, which effectively isolated the trial-by-trial link between interoception accuracy and confidence ratings.

#### Estimating Metacognition of Interoception with All Participants

As explained in the main text, we estimated metacognition about interoception using only those participants with a  $d'$  value greater than 0.25 (corresponding to roughly 54% correct) in the interoception task, comprising 45 out of 66 participants. We did this motivated by the principle that only in those participants would we even have the chance to observe and quantify metacognitive performance. Put another way, when someone’s behavioral accuracy is close to or at chance levels, it becomes increasingly difficult to quantify the degree to which they have metacognitive insight into the correctness of a given judgement. The inclusion of low-performing participants would mainly inject noise into the analytic process estimating metacognition about interoception.

Nevertheless, it was possible to replicate our main model with all participants ( $N = 66$ ) using identical fitting procedures as the model in the main text (i.e. 40,000 final samples, etc). This model returned  $R$ -hat values for the three group-level parameters capturing the mean and variance of the group-level distribution of individuals'  $M$ -ratio parameters, as well as the Actor's Bonus term that are all  $< 1.007$ . The mean sampled value of the "Actor's Bonus" parameter was very similar,  $M = 0.243$  (versus 0.257 from the  $N = 45$  analysis). While the 95% credible interval included zero (-0.0134, 0.503), 96.9% of samples were greater than zero (versus 98.1% from the  $N = 45$  analysis), suggesting a comparably sized and consistent effect, if marginally weakened by the inclusion of the 21 participants with very low levels of interoceptive performance on which basis to estimate metacognition about interoception.

### Highest Density Intervals

The credible intervals used in the main text analysis of MCMC samples are "equal tailed intervals", in which the positive and negative tails are constrained to have the same number of samples. There is debate as to whether this is the most appropriate type of interval to use with MCMC analyses (Kruschke, 2015), with some recommending the use of Highest Density Intervals (HDIs) instead, in which the interval describes the portion of the distribution with the highest density (it has the desired number of samples in the smallest range). The HDI defines the region in which all values inside have a higher probability density than any values outside the interval.

HDI in our analyses replicate the patterns reported in the main text and above. In the main text analysis focused on those individuals who demonstrated interoceptive ability, the 95% HDI of the Actor's Bonus  $\beta$  term is [0.011, 0.503]. In the version of that analysis discussed here in the supplementary materials (see above) leveraging all participants (including those with low/chance levels of interoceptive accuracy), the 95% HDI of the Actor's Bonus  $\beta$  term is [-0.013, 0.503].

### Hyperpriors for MCMC Fitting Procedure

Group-level parameters governing metacognitive efficiency (M-ratio) had hyperpriors selected to be relatively uninformative. Note that JAGS, the MCMC software used here to fit the model, parameterizes normal distributions with a mean term and a precision term. Precision is simply  $1/\text{variance}$ , or  $1/(\text{standard deviation})^2$ , but for clarity and simplicity in the main text, we discussed our model (and the normal distributions used) using standard deviation. The hyperpriors we adopted for our group-level parameters were  $\mu_M$ : uniform distribution between 0 and 4,  $\mu_{\text{Precision}}$ : gamma distribution with shape 0.001 and rate 0.001, and  $\beta$ : normal distribution with mean 0 and precision 0.25. The hyperprior on the beta parameter embodies the assumption that differences in metacognitive efficiency between Actors and Controls may be either positive or negative (mean = 0) and of a range of possible magnitudes (precision = 0.25, corresponding to standard deviation = 2). The version of the model used here (Fleming, 2017) also used constant subject-level priors on confidence criteria. These were chosen to cover reasonable ranges of expected parameter values relatively evenly (and then some), but to otherwise be relatively uninformative, and thus to reduce their impact on the sampling estimation procedure.

### Interoceptive Tasks

In the current study, we use a synchronization-style heartbeat detection task (Eichler & Katkin, 1994). Synchronization tasks have a risk of false negatives (labeling someone a poor interoceptor when they are good) because the task requires integrating interoceptive (heartbeat) and exteroceptive (auditory) signals, and because the subjective perception of heartbeat-tone lag may depend on individuals' physiology (Brenner & Ring, 2016). These criticisms render the measure *conservative*: if someone appears to be a good interoceptor, we can be confident that they are indeed, and that we have not accidentally classified someone as a good interoceptor when they are not (minimizing false positives). The most popular alternative interoceptive task, heartbeat counting (Schandry, 1981), has a significant risk of false positives (in addition to normal rates of false negatives). Participants can appear to be good interoceptors without being able to sense their heartbeat, simply by guessing, whether those guesses are educated or lucky. Because synchronization tasks are conservative, our results are thus more likely to underestimate than overestimate true interoceptive effect sizes (Brenner & Ring, 2016), and analyses focused on people classified by this task as good interoceptors (like our main metacognitive analysis) are unlikely to include individuals who should not be there, though some individuals may be errantly excluded.

Unlike in other perceptual modalities (e.g. vision), it has until very recently proven unfeasible to collect data on interoceptive discrimination over large numbers of trials. For example, as referenced in the main text, one previous study investigating interoception and metacognition (Garfinkel et al., 2015) used a 15-trial interoceptive task (note that this kind of design has been very common in this literature; we focus on this paper only for the purposes of a concrete example). When simulating the performance of 100,000 participants with good ground-truth

interoceptive ability ( $p(\text{correct}) = 0.7$ ) over 15 trials, we found the 95% CI of participants' observed  $p(\text{correct})$  was [0.47 0.93], illustrating that participants with identical moderately good 'true' ability would produce widely variable estimates of ability from chance levels to nearly perfect. In contrast, the 95% CI for the same participants in our study (with 160 trials) would be [0.63 0.77], a significant improvement and, notably, one that allows us to reliably categorize good interoceptors as performing substantially better than chance.

Low trial numbers also have consequences for the measurement of metacognitive awareness, as metacognitive sensitivity is a second-order statistic that relies on effective coverage of the confidence-accuracy matrix (see Fleming, 2017, for a discussion of biases introduced by low trial numbers in analyses of metacognition), with the result that more than 100 trials per participant is often desired for the most robust recovery of estimates. Some of these limitations were addressed in subsequent and recent research (Harrison et al, 2021), which estimated metacognitive abilities in respiration interoception with up to 60 trials per participant. However, these bounds are of course sensitive to the number of participants and the complexity of the inference sought. Here, for example, we sought to distinguish between two groups of participants, and because of limitations in participant recruitment, were unable to collect a very large sample – in this kind of setting, it becomes imperative to not just recover group-level metacognitive efficiency metrics, but robust individual-level metrics as well, and compensate for slightly lower participant numbers with increased numbers of trials. Finally, note that there have been some exciting recent developments in interoception tasks (e.g. Palmer, Ainley, & Tsakiris, 2019; Harrison et al, 2021) that suggest that the analytical and statistical sophistication of the field of interoception research is significantly increasing.

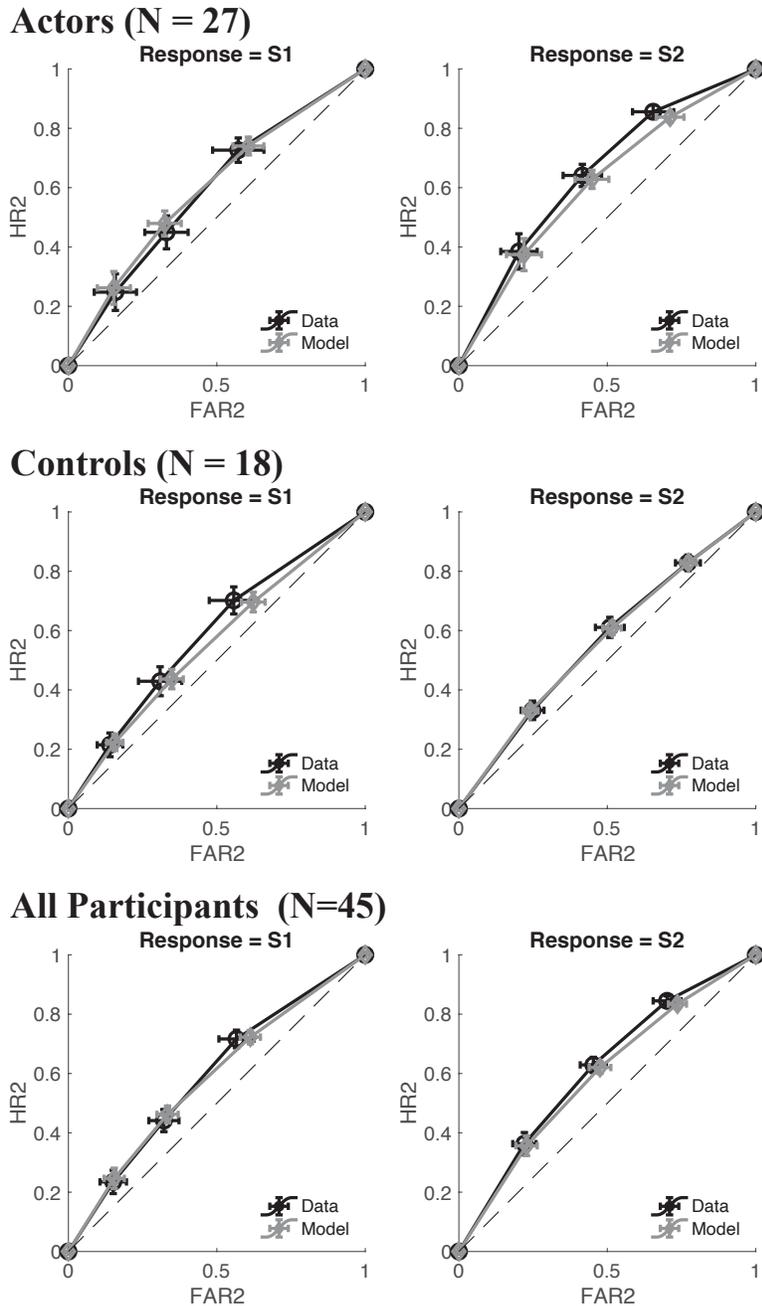


Figure S1: Plot of fitted/observed Type 2 (metacognitive) receiver-operating-characteristic (ROC) curves to visualize the fit of the model of metacognition used in the main analysis.

Overlap of the fitted Type 2 ROC curve values with those from the data are one indication that the model fit the data well. The top pair of plots represent the data and model fit for Actor participants, the middle pair for Control participants, and the bottom pair for all participants

## Interoception and metacognition in actors and non-actors: Supplementary Material

(though note all model predictions are extracted from the model used in the main text, which fit all participants simultaneously). HR2 (the y-axis) indicates the hit rate, FAR2 (the x-axis) indicates the false alarm rate, and S1/S2 refer to the two possible responses (in-sync or delayed, in reference to the cardiac interoception task). Points are means, and error bars are 95% CIs.

**Table S1: Summary of Participant Groups.**

Category	N	M/F	Age (years)	Height (inches)	Weight (lbs)	BMI (kg/m <sup>2</sup> )	STAIS	STAIT	Acting Experience Level
Tisch Actors	26	14/12	27.2 (3.0)	69.4 (3.3)	151.1 (25.0)	22.0 (2.6)	36.8 (10.4)	43.9 (11.3)	2.98 (0.1)
Other Actors	11	2/9	22.0 (3.1)	65.1 (3.8)	139.5 (27.2)	23.1 (4.0)	39.0 (8.0)	48.8 (9.5)	2.50 (0.5)
Non-Actor Controls	29	6/23	24.2 (5.1)	65.5 (2.7)	139.6 (29.1)	22.9 (4.6)	33.5 (7.7)	39.7 (10.8)	0.22 (0.4)
Tisch vs. Non-Actor Controls		p = 0.011	p = 0.012	p = 1.5e-5	p = 0.13	p = 0.40	p = 0.18	p = 0.17	p = 4.7e-36
Actors (Tisch & Other) vs. Non-Actor Controls		p = 0.054	p = 0.19	p = 0.0034	p = 0.24	p = 0.58	p = 0.07	p = 0.04	p = 1.3e-36

Summary of Participant Groups. Values reflect means with standard deviations in parentheses. STAIS and STAIT scores are each on a 20-80 scale, Acting Experience Level scores are on a 0-3 scale. P-values reflect the significance of chi-square tests for male/female ratios, and two-sample t-tests in all other cases between the Tisch actors and non-actor controls (first row of p-values) and all actors (collapsing across Tisch and other actors) and non-actor controls (second row of p-values).