Comparing normative and descriptive models of probability weighting

Benjamin Skjold*[1,2,3], Tobias S. Andersen†[2], and Oliver J. Hulme‡[1]

[1] *Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Amager & Hvidovre, Kettegard Allé 30, 2650, Hvidovre, Denmark.*
[2] *Section for Cognitive Systems, Department of Applied Mathematics and Computer Science, Building 321, Technical University of Denmark, 2800 Lyngby, Denmark*
[3] *London Mathematical Laboratory, 8 Margravine Gardens, W6 8RH London, UK*

October 18, 2021

Abstract

Probability weighting describes a systematic discrepancy between subjective probabilities and the objective probabilities of a given generative process. Since its introduction as part of prospect theory, probability weighting has been considered an irrational cognitive bias — an error of judgment committed by the decision maker. Recently, theories seeking to provide normative accounts for probability weighting have emerged. These attempt to explain why decision makers should express these discrepancies. One such theory, which we term the mechanistic model, proposes that probability weighting can be explained simply as a decision maker having greater uncertainty about the world than the experimenter. Such uncertainty arises naturally in decisions from experience, where the decision maker is expected to experience rare outcomes less often than the objective probabilities in the experimenter's code prescribe. Arcording to this theory, uncertainty decreases with increased experience, introducing a time-dependency whereby the discrepancy between subjective and objective probabilities narrows with time. This approach contrasts with prospect theory, a descriptive theory, which provides a functional form for describing the probability weighting, and treats the phenomena as a static trait of the decision maker. The two theoretical approaches thus make different experimental predictions. Here we present a Bayesian cognitive model that formalizes the two competing models, incorporating both models into one hierarchical latent mixture model. We explore how the different theories behave when faced with a simple experimental paradigm of decision making from experience. We simulate choice data from synthetic agents behaving in accordance with each theory, showing how this leads to discrepant experimental predictions that diverge as

* ✉ b.skjold@lml.org.uk
† ✉ toban@dtu.dk
‡ ✉ oliverh@drcmr.dk

the agents acquire more experience. We show that the same modeling framework when applied to the synthetic choice data, can accurately recover the correct model and that it can recover ground truth parameters, wherever they occur. This work can be taken as preregistration of the formal hypotheses of the two models for a simple experimental paradigm and provides the code necessary for future experimental implementation, model comparison, and parameter inference. Together, this demonstrates the tractability of empirically discriminating between normative and descriptive theories of probability weighting.

Keywords: Decision theory, probability weighting, prospect theory, Bayesian modeling

Introduction

The concept of probability weighting originated in prospect theory (Kahneman and Tversky 1979, PT)[1] and was later elaborated upon in cumulative prospect theory (Tversky and Kahneman 1992, CPT). The general idea of CPT is that it is not only the outcomes of a decision that are transformed non-linearly but also the probabilities that attach to those outcomes. Thus, CPT assumes that human decision making maximizes a utility function, which for a gamble with outcomes x and corresponding probabilities p is given by:

$$U = \Sigma v(x_i) w(p_i), \tag{1}$$

where $v(x)$ represents the subjective value function (relative to a reference point), and $w(p)$ represents the subjective probability weighting.

Several functional forms of the probability weighting function have been proposed. Here, we only consider the two-parameter function originally proposed by Goldstein and Einhorn (1987) (for others, see e.g. Tversky and Kahneman (1992) or Prelec (1998)):

$$w(p; \delta, \gamma) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}, \tag{2}$$

where γ is a *probability-sensitivity* parameter that primarily controls the curvature and δ is an *elevation* parameter that controls the placement of the inflection point, which has been interpreted as an indicator of the decision makers optimism regarding gambles (Tversky and Fox 1995). A visual representation with different parameter settings can be seen in Fig. 1.

After the introduction of PT and CPT, many experiments were conducted to estimate the parameters of the probability weighting function and its sibling, the value function. The general setup in most experiments has been to task the agent with choosing between two prospects: a risky gamble composed of monetary outcomes and stated probabilities, versus a sure gain of a monetary outcome (see Camerer and Ho (1994); Gonzalez and Wu (1996, 1999); Tversky and Fox (1995); Tversky and Kahneman (1992), see also Birnbaum and McIntosh (1996) for a different approach). This is known as decision making from description because the prospects are described explicitly to the decision maker (Hertwig et al. 2004). Although the estimated parameter estimates varied across these studies, the general results are similar. Namely, that unlikely events are overweighted and likely events underweighted; resulting in the famous inverse-S shape (see Fig. 1, red/dashed line). However, within the last decades, researchers have been increasingly interested in probabilities that are derived from experience rather than from description. In many cases, this is considered to better resemble choices made outside the laboratory, where "people often must make choices without a description of possible choice outcomes, let alone the probabilities that attach to them." (Hertwig et al. 2004). It came as a surprise to many in the field when several studies revealed observed choices, which indicated not overweighting of rare events, but underweighting (Barkan et al. 1998; Barron and Erev 2003; Weber et al. 2004) - as visualized by the blue/dash-dotted line in Fig 1.

[1]We note that there is literature dating back to the late 1940s, which consider transformations of probability within decision science - see e.g. Preston and Baratta (1948). However, we will not comment further on this literature, as its theoretical foundation differs from that of the modern theories.

Hertwig et al. (2004) formulates this with the same four-fold pattern introduced by Kahneman and Tversky (1979); Tversky and Kahneman (1992), proposing that for decisions from experience rare events are underweighted rather than overweighted. This discrepancy between decisions from description and decisions from experience is known as the *description-experience gap* (Hertwig and Erev 2009). A meta-analysis conducted by Wulff et al. (2018), including nearly 70.000 decisions from experience, report that the description-experience gap is a robust finding. However, in line with the criticism of the description-experience gap (see e.g. Fox and Hadar (2006) and Rakow et al. (2008)), Wulff et al. show that limitations of the experiments, such as reliance on few trials is an important contributor to the gap. The smaller the number of trials within each experiment the larger the possible sampling error, where the frequencies of events deviate from their objective probabilities, especially for lower and higher probability events. Further, if the agent's decision making influences the information they receive about the generative process, then this can lead to biases that may contribute to an apparent gap between the subjective and the objective probabilities. In other words, in some paradigms, the epistemic value of obtaining more experience with a given process may confound the interpretation of the choice behavior. Lastly, Wulff et al. show that the overall pattern of the probability weighting emerging from decisions from experience can be mixed: Hau et al. (2008), Camilleri and Newell (2009), Ungemach et al. (2009) and Hau et al. (2010), for example, report regular-S shaped functions (blue/dash-dotted line in Fig. 1), while Abdellaoui et al. (2011), Glöckner et al. (2016) and Kellen et al. (2016) report inverse-S (red/dashed line in Fig. 1). The discrepancies have so far been attributed to methodological differences; for example, the use of the sampling, partial feedback, or full feedback paradigm (see Hertwig and Erev (2009) for a description of these paradigms). Even within these paradigms, there can be discrepancies attributed to the probabilities of the rare event, as well as the order in which the gambles are presented, and whether the gambles are gains only, loss only, or mixed (Wulff et al. 2018).

Recently, there has been increasing effort to generate normative theories of probability weighting. Here we focus primarily on one of these: Peters et al. (2020), but other approaches have also been proposed (see e.g. Stewart et al. (2006) or Steiner and Stewart (2016)). Peters et al. propose that "[...] probability weighting means that a decision maker has greater uncertainty about the world than the observer". They consider a decision maker, such as a subject in a study, and a disinterested observer (observer, henceforth), such as an experimenter observing the behavior of the subject. The decision maker has greater uncertainty within a decision from experience experiment because they have to base their probability estimates on the frequencies of events they have observed, and each of these frequencies is compatible with a distribution of different probabilities. Further, they are basing this estimation on their memory of the frequencies and calculating probabilities with finite computational precision, both of which generate uncertainty as to what the underlying objective probability is. Where the observer is an experimenter, they will typically know the generative process exactly because they have specified this by design. Hence there is typically a fundamental discrepancy in the uncertainties of the decision maker and the observer when it comes to estimating probability.

In decision making from experience, for events with low objective probabilities, the relative frequencies observed are expected to be lower than the objective probability. Symmetrically, for events with high objective probabilities, the relative frequencies observed are expected to be higher than the objective probability. The reason for this is as follows.

For an asymptotic probability density $p(x)$, the number of events $n(x)$ that is experienced in an interval of T observations, is proportional to $p(x)$ and T. In a simple Poissonian case where the probabilities are constant, the count variable $n(x)$ is a random variable whose uncertainty scales with $\sqrt{n(x)}$. Peters et al. define an estimate for $p(x)$ as:

$$\hat{p}(x) \equiv \frac{n(x)}{T}, \tag{3}$$

and its uncertainty is the standard error of this estimate:

$$\varepsilon[\hat{p}(x)] \equiv \sqrt{\frac{\hat{p}(x)}{T}} = \frac{\sqrt{n(x)}}{T}. \tag{4}$$

Note that this standard error shrinks as the probability decreases. However, the relative error in the estimate is $\frac{1}{\hat{p}(x)T}$ which grows as the event becomes rarer. This is the reason why low probabilities come with larger relative errors.

In the mechanistic model, the decision maker estimates their subjective probability according to the following equation:

$$w(x) = \frac{\hat{p}(x)+\varepsilon[\hat{p}(x)]}{\int_{-\infty}^{\infty} \hat{p}(s)+\varepsilon[\hat{p}(s)]ds} \tag{5a}$$

$$= \frac{\frac{n(x)}{T}+\frac{\sqrt{n(x)}}{T}}{\int_{-\infty}^{\infty} \frac{n(s)}{T}+\frac{\sqrt{n(s)}}{T}ds}, \tag{5b}$$

where the denominator is a normalizing factor ensuring $w \in [0, 1]$. The decision maker is in effect estimating $\hat{p}(x)$ adding the standard error of this estimate and normalising. The motivation behind this is to approximate a reasonable worst case estimate of $p(x)$.

An important observation in the mechanistic model is that the error term depends on T and thus decreases with the number of observations. As T grows, $w(x)$ therefore approaches the asymptotic probability density, $p(x)$. In contrast to CPT, this mechanistic model is time dependent where the error, decreases with increasing information, and in the theoretical case of infinite sampling and perfect memory, the weighting function will be linear (/no weighting) - illustrated in Fig. 2.

The goal of this paper is to bring descriptive models and normative models together, drawing out their disagreements, and assessing how easily this can be experimentally tested. Specifically, we embed the CPT model and the mechanistic model within a single modeling framework that will allow us to explore via simulation how experimentally tractable the disagreements between the two are, and over what time scale these emerge. This can be taken as a formal pre-registration of hypotheses, and as preparatory work for experimental implementation, providing all code necessary for simulation and inference.

Methods

Experimental design. In our simulations we use the following decision making from experience paradigm (Fig. 3). On each trial, the agents face the choice between two

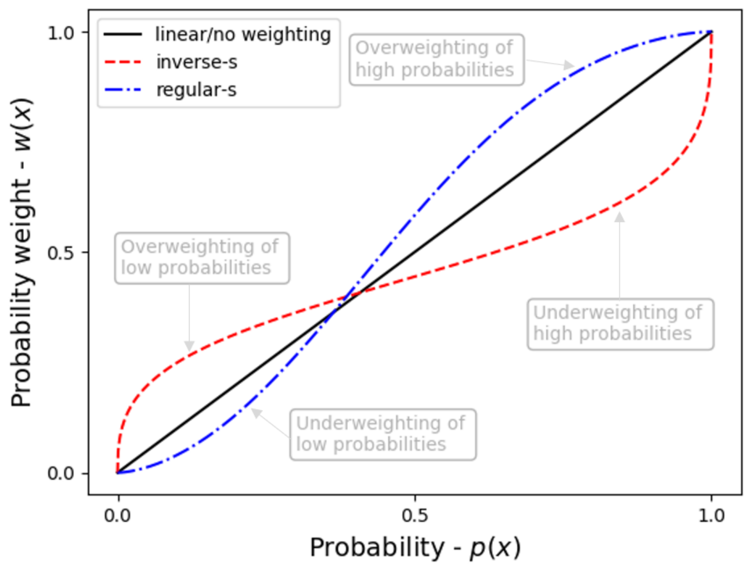


Fig. 1. Different shapes of the probability weighting function: red line (dash) corresponding to inverse-S (eq. 2; $\delta = 0.8, \gamma = 0.4$), black line (solid) corresponding to linear (equivalent to no weighting) and blue line (dash-dot) corresponding to regular-S shape (eq. 2; $\delta = 1.4, \gamma = 1.6$).

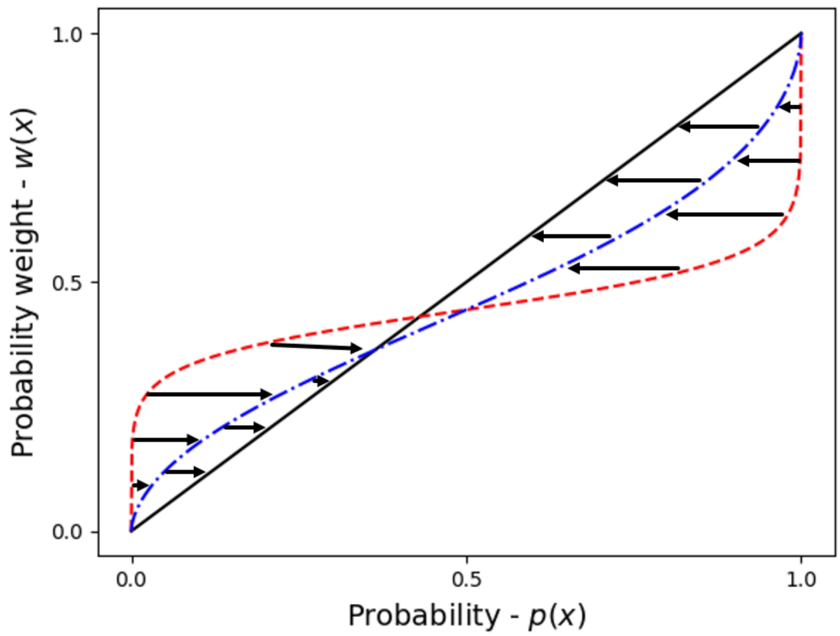


Fig. 2. Probability weighting function as predicted by th mechanistic model for increasing number of observations: The PWF plotted for different values of T; the red/dashed line representing small T, the blue/dashed-dotted line representing a larger T and the black/solid line representing (the theoretically) perfect information ($T = \infty$).

gambles, A and B, which have different payoffs, each with different fixed probabilities. The payoffs and probabilities are all unknown at the start of the experiment, and can only be inferred via experience. Upon choosing one of the gambles, the agent receives the monetary outcome associated with that gamble, which is added to their current wealth. Agents are informed of their current wealth at all times. They also observe the outcome of the unchosen gamble which does not add to their current wealth. The agents thus receive full feedback about the outcomes of both gambles, even though it is only the chosen gamble that impacts their wealth. Their new wealth carries over into the next trial where they face the same choice between A and B. The advantage of providing full feedback from both chosen and unchosen gambles is that it makes the information that the agent has obtained independent of their choices.

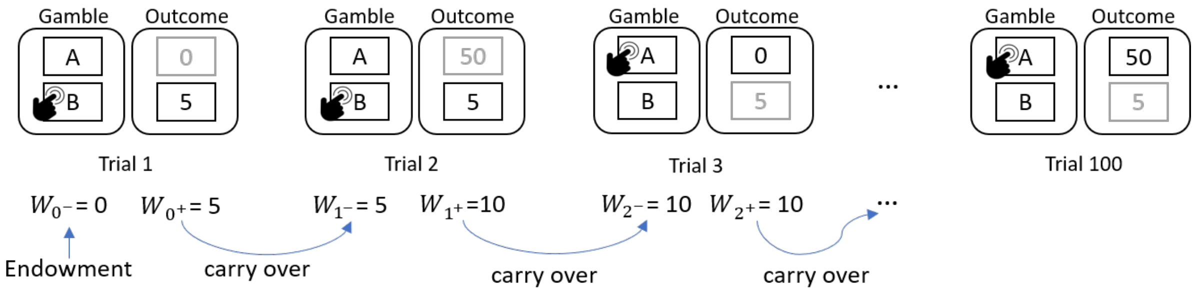


Fig. 3. Experimental design: An example sequence of 100 choices for an underlying gamble - in the example illustrated they choose between A which yields 50 with probability 0.1, 0 otherwise, and B which yields 5 with probability 1 (gamble set 1, see Tab. 1). The agent must choose between gamble A and B in each trial, and is presented with the outcomes for both gambles after choosing. Below the sequence the agent's wealth is illustrated, with an endowment of 0, wealth updated within each trial according to the agent's choice, and wealth carries over between trials.

Gambles. We consider 24 qualitatively different gamble pairs. A gamble pair consists of two gambles for the decision maker to choose between. Gamble A is a conventional gamble consisting of two (monetary) outcomes x_1, x_2 with corresponding underlying probabilities $p_1, 1-p_1$ - written in the form $(x_1, p_1; x_2)$. Gamble B, on the other hand, (with an abuse of nomenclature) is a sure gain, set to the expected value (EV) of the gamble it is paired with, rounded to the nearest integer (see example in Fig. 3 gamble B). We illustrate the 24 gamble pairs that we consider in Tab. 1. We have chosen these gamble pairs, as the gambles are the gains only gambles used when developing the CPT-model - see Tversky and Kahneman (1992, Tab. 3).[2]

For each of the gamble pairs, the gambles are transformed into a sequence of gambles of length $n_T = 100$; for each trial (n_T times) drawing a (pseudo-)random number from a uniform distribution to generate an outcome for gamble A, G_A. Note that as the outcome of gamble B is set to be the EV of G_A and is therefore just a series of $EV[G_A]$ of length n_T. To ensure the underlying distributions are well represented by the sequence, we check that the difference between underlying probability and the frequency distribution is within a threshold set to 0.01. We repeat this for each agent, such that all agents face different sequence with the same underlying probability distribution.

Our simulations will model how three different types of synthetic agents behave according to the two theories when encountering this paradigm. We refer to these types of agents as different species (see Fig. 5.a). Where the model allows for individual variation we refer to these as different phenotypes. Across different agents within a species, the

[2]except the gambles where $p_1 = 0.01$ or $p_1 = 0.99$, as these resulted in relative errors between the frequency distribution and underlying probability distribution outside an acceptable range.

sequence of gamble pairs and outcomes are randomized. However, across species the sequence of gamble pairs and outcomes are for agent i is identical, but because the decisions can be different, the wealth trajectories can be different.

	Probability								
Outcomes	~~.01~~	.05	.10	.25	.50	.75	.90	.95	~~.99~~
(0,50)			5		25		45		
(0,100)		5		25	50	75		95	
(0,200)	~~2~~		20		100		180		~~198~~
~~(0,400)~~	~~4~~								~~396~~
(50,100)			55		75		95		
(50,150)		55		75	100	125		145	
(100,200)		105		125	150	175		195	

Tab. 1. Gamble pairs: The body of the table contains the sure value (Gamble B) for the gamble (Gamble A) of each gamble pair. The two possible outcomes of each gamble A are given in the left-hand column; the probability of the second (i.e. higher) outcome of gamble A is given by its column. For example, the value of 5 in the upper left corner is the outcome for gamble B corresponding to (50, 0.1;0) as gamble A. Note these are retrieved from Tversky and Kahneman (1992) - crossed out gambles are the four gambles that are left out.

Model space. We describe the following models by specifying the probability weighting function, as well as a stochastic choice function which accommodates stochasticity in the agents choice when choosing between two gambles of differing utilities Peters et al. (2020); Stott (2006).[3]

Cumulative Prospect Theory (CPT) weighting function as presented in eq. 2 with two parameters:

$$w(x;\delta,\gamma) = \frac{\delta\hat{p}(x)^{\gamma}}{\delta\hat{p}(x)^{\gamma} + (1-\hat{p}(x))^{\gamma}}, \tag{6}$$

where γ is a probability-sensitivity parameter determining the curvature and δ an elevation parameter determining the elevation of the inflection point.

Mechanistic-model weighting function as presented in eq. 5:

$$w(x) = \frac{\hat{p}(x) + \varepsilon[\hat{p}(x)]}{\hat{p}(x) + \varepsilon[\hat{p}(x)] + (1-\hat{p}(x) + \varepsilon[1-\hat{p}(x)])}. \tag{7}$$

Note that this probability weighting function does not have any latent parameters.

Expected value. The expected value as presented in eq. 1 (with linear utility) for each gamble (done identically for each of the two models):

$$\langle x^A \rangle = w(\hat{p}_1^A) \cdot x_1^A + w(1-\hat{p}_1^A) \cdot x_2^A, \tag{8}$$

and equivalent for gamble B.

[3] We assume linear utility to focus solely on probability weighting and note that within the literature it is argued that the probability weighting is qualitatively unaffected by the utility function (Tversky and Kahneman 1992).

We denote the difference between the gamble A and gamble B by Δ and thus write the difference for each choice as:

$$\langle x\rangle^{\Delta} = \langle x^{A}\rangle - \langle x^{B}\rangle. \tag{9}$$

Stochastic choice function. The stochastic choice function is likewise identical for the two models and is comprised of a logistic function with sensitivity parameter β:

$$\theta\left(\langle x\rangle^{\Delta}\right) = \frac{1}{1+e^{-\beta\langle x\rangle^{\Delta}}}, \tag{10}$$

where θ evaluates to the probability of choosing gamble A.

Sampling procedure. We estimate and simulate the latent parameters of each model by hierarchical Bayesian cognitive modeling (Lee and Wagenmakers 2013). Via its hierarchical structure, we can inform estimations of individuals from the group, constraining extreme values that might arise due to uncertainties (Nilsson et al. 2011). We sample from the posterior distributions with Monte-Carlo Markov Chain (MCMC), sampling via JAGS (v4.03), called from MATLAB™ (v9.5.0.944444 (R2018b), Mathworks®, `mathworks.com`) via the interface MATJAGS (v1.3, `psiexp.ss.uci.edu/research/programs\_data/jags`). For all simulations (i.e. simulating choices, model recovery, and parameter recovery), we estimate the two models using a hierarchical latent mixture (HLM) model. In Fig. 4 we illustrate the HLM model via the standard graphical representation of Bayesian cognitive models. We list the distributional and structural equations in Fig 4b-d. These depend on whether they are used for simulating choices, model recovery, or parameter recovery and are thus described again in the respective sections below. We note that CPT- and the mechanistic model are technically only sub-models of the single HLM model. However, we refer to them simply as models for simplicity and consistency. For simulating choices, we use just one chain and one sample resulting in one choice per agent per simulated trial, where we, for model recovery and parameter recovery, use four independent chains, 10^3 samples per chain, and include a burn-in > 500. Lastly, we establish convergence via monitoring R-hat values between 0.99 to 1.01 (Brooks et al. 2011; Gelman et al. 2014).

Simulating choices We simulate choices from synthetic agents in order to evaluate whether the model estimation methods are capable of identifying the data generating model (the species) as well as its parameters (its phenotype). These agents come from group level distributions, for which 'ground truth' parameter values are set a priori. Note that where it is stated we use Dirac delta functions, these are approximated by very narrow intervals on uniform distributions.

As previously stated, we simulate three different species, each containing 10 agents; two species of the CPT-model, where one expresses the inverse-S weighting function and the other regular-S, and one species of the mechanistic model (see Fig. 5.a). We obtain these simulated choices by running the HLM model three times with no choice data as input variables, and by setting the prior distribution on the indicator variable (which selects each model), Z, to $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, respectively (see Fig. 4d).[4] For

[4]Note that the only difference between the two species of CPT-model is the prior on the mean of the sensitivity parameter, γ. Thus, we treat them as one species for recovery.

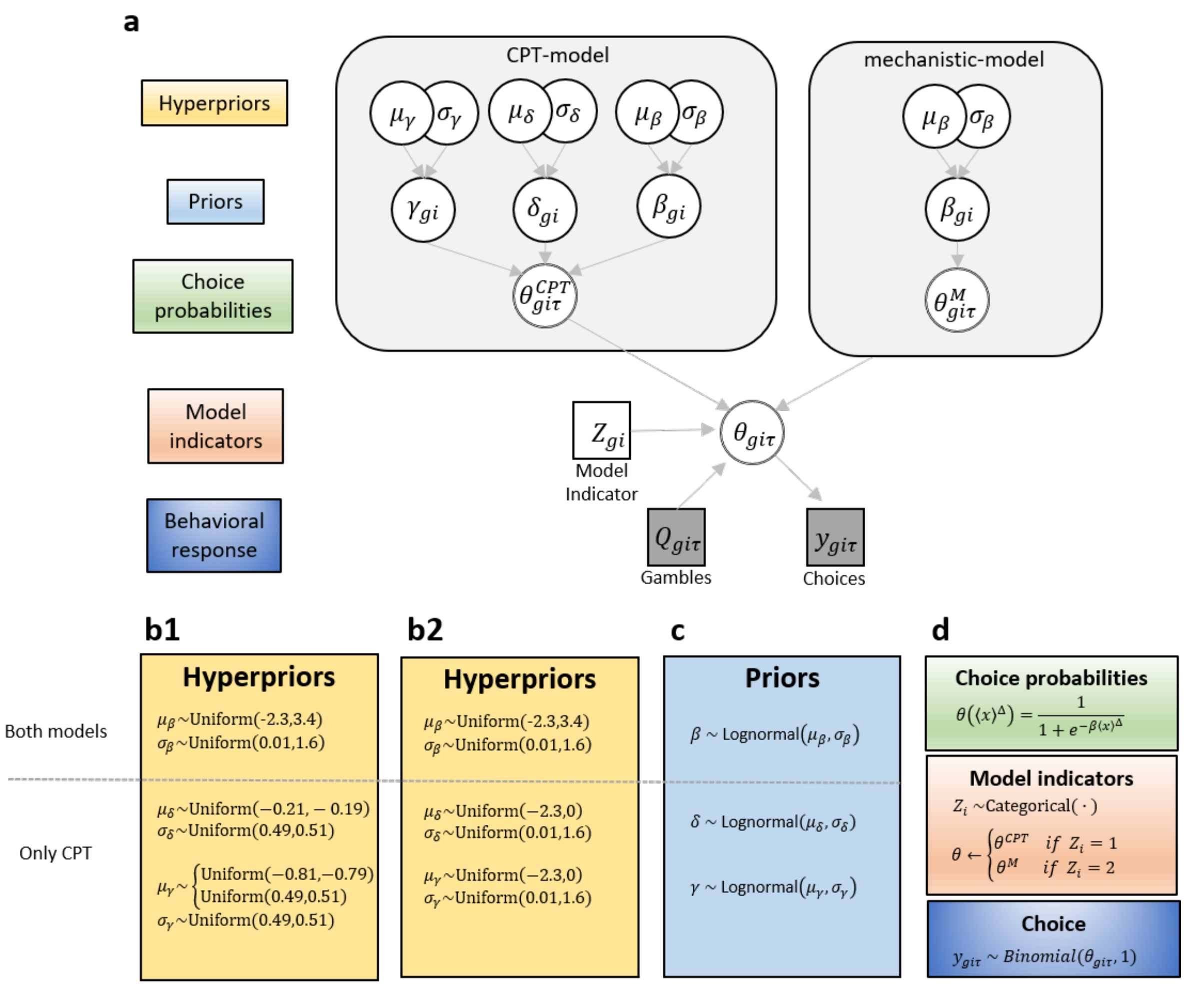


Fig. 4. Hierarchical latent mixture model. a, graphical representation of hierarchical Bayesian model for estimating latent mixtures for two different models. Circular nodes denote continuous variables, square nodes discrete variables; shaded nodes denote observed variables, unshaded nodes unobserved variables; single bordered nodes denote stochastic variables, double bordered nodes deterministic variables. Labels on the left describe the role of each parameter in the model. **b1,** hyperprior distributions used when simulating choices; interval in uniform distributions set very narrow, thus acting as an approximation to a Dirac delta functions. **b2,** hyperprior distributions used for model- and parameter recovery. **c,** prior distributions for both simulating choices, model recovery, and parameter recovery. **d,** choice functions and choice generating distributions.

the sensitivity parameter used in the stochastic choice function, we follow Nilsson et al. (2011) and assume the group mean for β to lie in the interval $(0, 30)$ and therefore set $\mu_\beta \sim \text{Uniform}(-2.3, 3.4)$ and the standard deviation to $\sigma \sim \text{Uniform}(0.01, 1.6)$. For the weighting function parameters (only in the CPT model), we fix the prior distributions such that for species one the mean values follow approximations from the literature providing an inverse-S shaped weighting function (Gonzalez and Wu 1996, 1999) and for species two we fix the mean to provide regular-S shaped weighting functions (Hau et al. 2008). Hence, we assume $\delta = 0.8$, $\gamma_1 = 0.5$ and $\gamma_2 = 1.5$, and therefore set $\mu_\delta = -0.2$, $\mu_{\gamma_1} = -0.8$, $\mu_{\gamma_2} = 0.5$ and further set $\sigma_\delta = \sigma_{\gamma_1} = \sigma_{\gamma_2} = 0.5$ - Fig. 4.b1.

Model and parameter recovery. We attempt both model and parameter recovery by estimating the HLM model with the previously simulated choices as input. For both, we follow Nilsson et al. (2011) and set weakly informed hyperpriors. As when simulating choices, we assume β to have common group level distribution across models, while δ and γ are only applied to the CPT-model. We note that all priors are non-negative and thus we assume each to come from uninformative log-normal distributions with uninformative priors on group mean, μ, and variance, σ^2, which are both drawn from uniform distributions. We assume group mean for β to lie in the interval $(0, 30)$ and therefore set $\mu_\beta \sim \text{Uniform}(-2.3, 3.4)$ and the standard deviation to $\sigma_\beta \sim \text{Uniform}(0.01, 1.6)$. For both the elevation parameter, δ, and sensitivity parameter, γ, we assume the group means to lie in the interval $(0, 2)$; $\mu_\delta \sim \mu_\gamma \sim \text{Uniform}(-2.3, 0)$, $\sigma_\delta \sim \sigma_\gamma \sim \text{Uniform}(0.01, 1.6)$ - Fig. 4.b2.

We perform parameter recovery separately for each of the three species. Since the CPT model is descriptive rather than normative, where is no a priori specification for what value the probability weighting parameters should take on. As such we specify uniform priors over its parameter space, which results in probability weighting functions that span from regular-S shape to inverse-S shapes (see Fig. 1).

The CPT model is used to estimate the parameters for the choice data generated by both the mechanistic model and the CPT models. This is appropriate because the CPT model is descriptive enough to approximate the probability weighting functions generated by the mechanistic model. The parameter values estimated for the mechanistic model are simply an approximation of what CPT model fits would look like when faced with an agent from the mechanistic model. As such the parameters estimated via CPT model can be informative of whether the mechanistic model is a good model of the data.

To evaluate the potential evolution of the weighting function over time, we estimate the parameters for three experiments of various lengths. We refer to these as chunks 1-3, such chunk 3 is the full experiment of $N_T = 100$ trials, chunk 2 is a nested subset consisting of the first 50 trials, and chunk 1 is a further subset consisting of only the first 10 trials. By estimating the parameters independently for each chunk, we estimate parameters for a various number of trials and thus we are able to evaluate the probability weighting function over time.

We performed model recovery by scrutinizing the posterior distribution of the indicator variables which controls the latent mixtures of the two models (Fig 4). This allows comparison between the qualitatively different models with the HLM model, by integrating over all parameters in each model (with respect to the respective priors) and penalizing for increasing model complexity (Lee and Wagenmakers 2013). For model recovery, the indicator variable is set with uninformed uniform priors but is free to vary for each agent. We calculate the likelihood for an agent's choices given their species i with Bayes' theorem: $p(Z = i|D) \propto p(D|Z = i)p(Z = i)$. As the indicator variable is set to be uniform, i.e. $p(Z = i) = p(Z = j)$, the posterior distribution of the indicator variable gives the marginal likelihood of each model, and from which Bayes factors can be directly computed. Finally, we estimate the posterior model probabilities of each species, and their associated and Bayes factors via the Variational Bayesian Analysis toolbox (Daunizeau et al. 2014, `mbb-team.github.io/VBA-toolbox/`).

Data and code availability. All datasets and code needed to replicate the findings presented here are available in the 'Comparing-normative-and-descriptive-models-of-probability-weighting ' repository: `https://github.com/benj1003/Comparing-normative-and-descriptive-models-of-probability-weighting`.

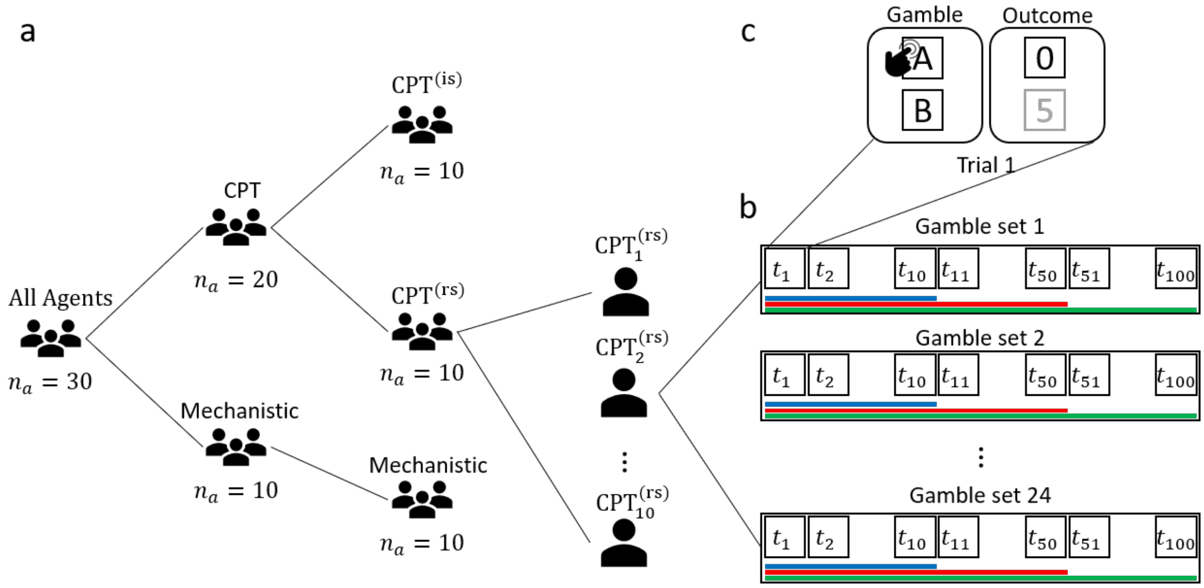


Fig. 5. Methods overview: a, illustration of the division of agents; from left illustrating all agents, which are split into two groups that behave according to the two theories. The CPT group is further split into two groups - CPT$^{(is)}$ and CPT$^{(rs)}$, where CPT$^{(is)}$ agents have priors on the parameters such that we expect inverse-s shaped probability weighting function and (CPT$^{rs)}$ such that we expect regular-s shaped probability weighting function. Thus, in the analysis we consider three different types of agents, which each consist of agents. **b,** each synthetic agent is presented with 24 different gamble sets, which each consist of 100 trials. For the analysis we split the trials into three nested chunks (referred to as chunk 1-3), where chunk 1 is the first 10 trials, as indicated by the blue bar, chunk 2 is the first 50 trials (including the first 10), as indicated by the red bar and chunk 3 is all 100 trials, as indicated by the red bar. **c,** each trial consist of a gamble screen, from which the agent must choose between gamble A or gamble B, and an outcome screen that is presented to the agent following the choice, which presents the outcome associated with both gambles in that trial.

Results

Parameter recovery. Initially, we investigated if the HLM model was able to recover the 'ground truth'/predicted parameters for both models. For the CPT-agents, we know the 'ground truth' parameters for all agents, as we simulate the choices for the agents based on the same parameters; δ_g, γ_g. We, therefore, examined the correlation between the 'ground truth' parameter and the corresponding maximum a posteriori probability (MAP) estimates; δ_e, γ_e. We illustrate this in Fig. 6 with a scatter plot of the two weighting function parameters for each of the 3 chunks. Here, it is evident that the MAP estimates and the 'ground truth' parameters are strongly correlated (Pearson correlation coefficient above 0.99 for all chunks) , which suggests that the HLM model can accurately recover the parameters for the CPT-agents. Whilst this may be trivial, it is nevertheless important to demonstrate for confidence in the HLM model.

An alternative way to express the estimation error is to investigate the difference between the 'ground truth'/predicted weighting function (which we denote $\bar{w}(x)$) and the estimated weighting function, $w(x)$. An advantage of this approach is that we can evaluate how well we recover the weighting functions for the mechanistic-agents, by evaluating against the predicted weighting function (5). Thus, we transform all estimated (and 'ground truth') parameters into a weighting function by eq. 2 and evaluate $\bar{w}(x) - w(x)$. We illustrate the average estimation error (including 95 % confidence interval) for all agents within each 'ground truth' model in Fig. 7. This shows that the error varies systematically with the estimated probability for both models, for which the origin, though, is unknown. However, the error in both models is of such low magnitude that we deem it effectively negligible for the pragmatic purpose of this paper and any experiment in which such parameters are estimated. We, therefore find that the HLM model is able to recover the 'ground truth' parameters for both the CPT and mechanistic models. In Appendix A.1 and A.2 we plot the weighting functions for all agents individually, which show that the model recovers the parameters with low estimation error for all agents and adds credence to the possibility of discriminating between these models via experiments similar to those simulated here.

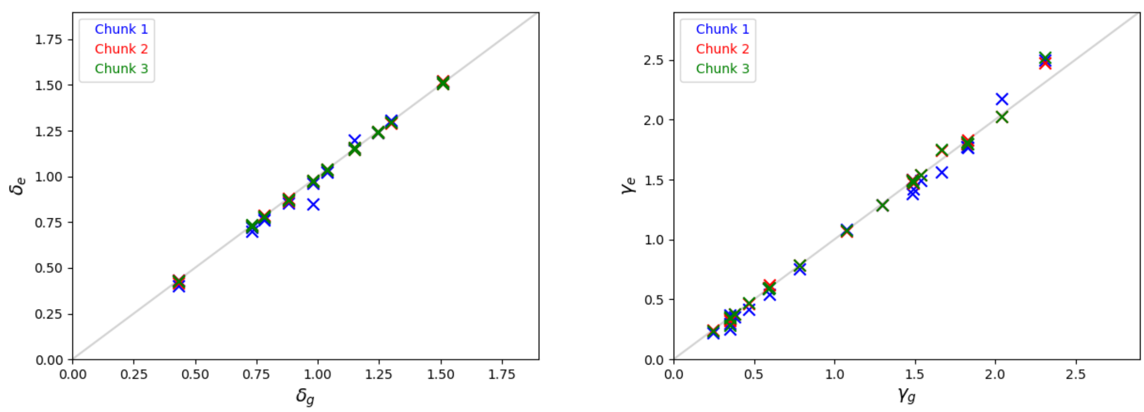


Fig. 6. Estimated parameters against 'ground truth' parameters for CPT-agents: Left: Estimated elevation parameter (δ) plotted against the 'ground truth' parameter. Right: Estimated sensitivity parameter (γ) plotted against the 'ground truth' parameter. Both: light-grey line representing perfect correlation; different chunks represented by different colors.

We next examined the estimated parameters in weighting function parameter space, i.e. γ_e against δ_e (2) - we illustrate this in Fig. 8 (Left panel: CPT-agents; Right panel: mechanistic-agents). Here, two important observations are made: 1) the CPT-model (left panel) predicts a variety of agents, while the mechanistic-model (right panel) has a very narrow predictive mass, and 2) when investigating the parameters for each agent, it is clear that for the CPT-agents (left panel zoom) there is no systematic bias in the parameters across the different chunks, while for the mechanistic-agents (right panel zoom) there is a clear tendency for the sensitivity parameter, γ, to increase towards 1, which is equivalent to 'moving' towards linear/no weighting, as expected - illustrated in Fig. 2. This shows that the framework provided is able to capture the differences in model structure between the two models, as well as the evolution of parameters and thus the probability weighting function over time.

Model recovery. Lastly, we compare the predictive accuracy of the two models on the basis of the simulated choices. MCMC sampling of the HLM model results in posterior

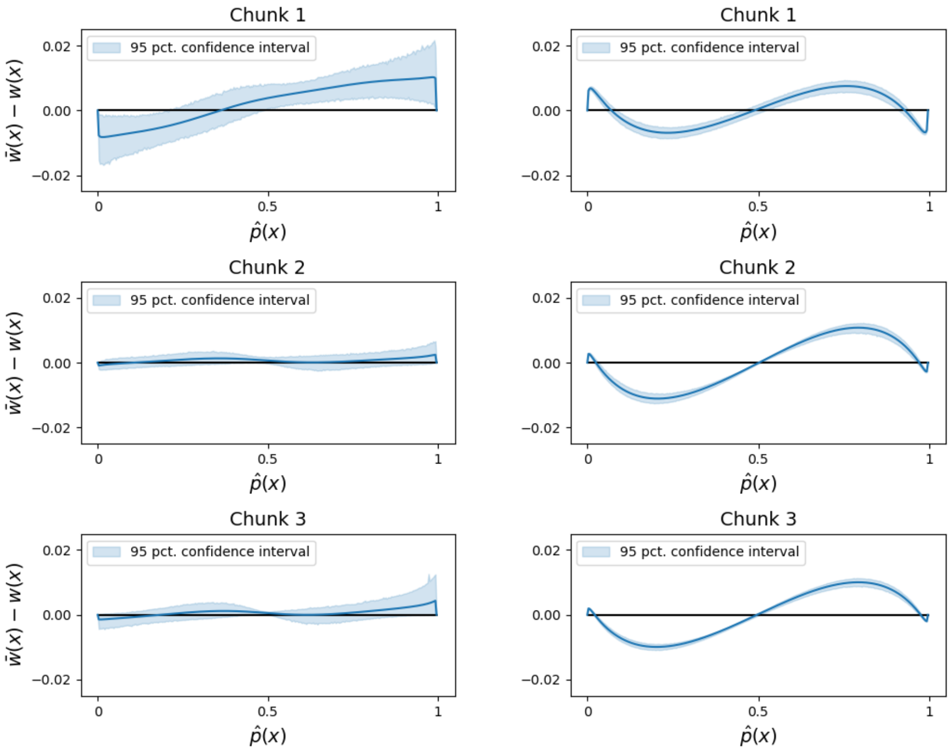


Fig. 7. Estimation error as a function of $\hat{p}(x)$: Left: Difference between estimated weighting function and 'ground truth' weighting function averaged over the 20 CPT-agents, including 95 % confidence interval. Right: Difference between estimated weighting function and predicted weighting function averaged over the 10 mechanistic-agents, including 95 % confidence interval.

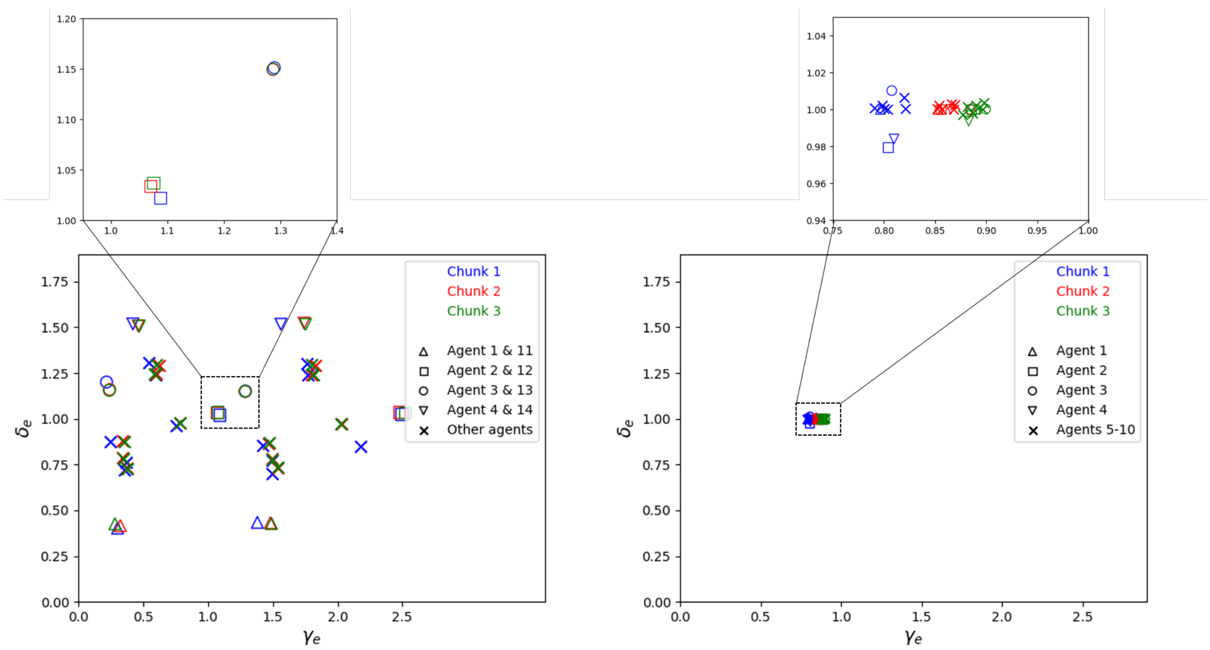


Fig. 8. Weighting function (eq. 2) parameter space for both models: Left: Estimated parameters for all 20 CPT-agents, showing a course of variety of agents being predicted; zoom showing no systematic bias across chunks for two arbitrary agents. Right: Estimated parameters for all 10 mechanistic-agents, showing very narrow predictive mass; zoom showing a clear systematic bias of $\gamma \to 1$ for $T \to \infty$ as expected, illustrated in Fig. 2.

frequencies for the model indicator variable, which one can interpret as posterior probabilities for each model, estimated for each agent. Evident from Fig. 9, all agents have all of their probability mass located over their 'ground truth' model ($BF_{\text{CPT-CPT}}$=41, $BF_{\text{mechanistic-mechanistic}}$=21). Thus, there is strong evidence that the framework provided correctly identifies the 'ground truth' model.

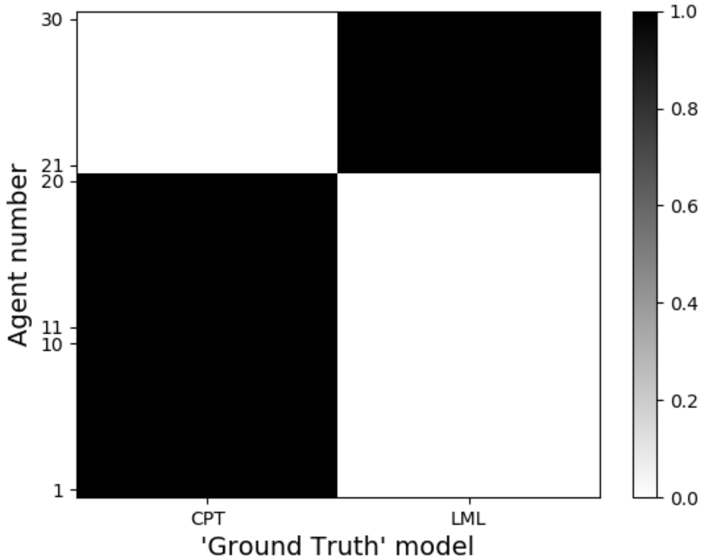


Fig. 9. Posterior model probabilities for each model based on the indicator variable; 'Ground truth' model for Agent 1-10 being CPT (inverse-S); Agent 11-20 CPT (regular-S); Agent 21-30 mechanistic.The black white axis reveals the posterior model probability for the model on the x-axis.

Discussion

In embedding both CPT and a mechanistic model of probability weighting within a single model, we demonstrate how a simple decision making from experience paradigm can reveal a disagreement between the two theories. Whilst CPT assumes the phenotype that is probability distortion static, the mechanistic model assumes a dynamic phenotype in which probability distortion arises from the uncertainty of experienced frequencies, which diminishes with the increasing certainty that attains from more experience. We show that on a short time scale of 90 trials, the mechanistic model diminishes the degree of probability weighting from an inverse-S shape to a form that is increasingly well approximated by a linear probability weighting function, which in effect culminates in no probability weighting. We show that on these short timescales model recovery and parameter recovery is easily attained and that the volume of data necessary is feasible for real experiements.

Cognitive plausibility. The mechanistic model was built as a null model, deliberately avoiding cognitive constraints that would make the model more plausible as a psychological model of human behavior. In its present form the mechanistic model (implausibly) assumes an infinite memory capacity, and infinitely precise tracking of observed frequencies. Furthermore there is no accounting for the possibility that the probabilities of the underlying generative processes could be non-stationary, as would be realistic in most natural settings. Subsequent modelling would of course add in more plausible cognitive constraints, allowing inference on their necessity in explaining experimental data.

Paradigm design. Much criticism in the decision from experience literature focuses on either sampling error or the lack of independence between trials. In terms of sampling error, the isssue is that the experienced probabilities, which the decision maker base their decision on, may differ from objective probabilities, of which the decision is often evaluated against. In terms of independence the issue is that the results are affected by other factors, such as the decicion maker being faced with exploration-exploitation trade-offs. In the experimental setup presented here, we eliminate these limitations by simulating and inferring weighting functions as a function of the observed frequencies, rather than the objective probabilities. Furthermore by using a full feedback paradigm, where the agents are presented with all outcomes after each trial, we attain independence between the agents choices and their experience.

Experimental constraints. We constrained the length of the gamble sequences ($n_T = 100$), such that we could test a wide range of underlying gambles ($n_g = 24$) and still provide simulated results that are compatible with the time scale and resources available to experimenters. Given this constraint we do not test extreme probability events that can not be reliable expressed within this timescale (for intuitive example of this issue we refer to Peters et al. 2020, p.10). This might be the reason for minor systematic error that is seen in Fig. 7. However, for the purpose of this paper, we argue that the potential increase in accuracy these gambles would add does not outweigh the cost associated with increased burden of data collection (unless of primary scientific interest).

Status of this paper This paper is unusual insofar as it prespecifies models and hypotheses in a way that effectively pre-registers them for future experimental test. In doing so, we have shown that it is possible to use principled Bayesian models of cognition to discriminate between descriptive and normative theories of subjective probability, laying a path to subsequent real experiments that can resolve this theoretical disagreement.

References

Abdellaoui, M., L'Haridon, O., and Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? Management Science, 57(10):1879–1895.

Barkan, R., Zohar, D., and Erev, I. (1998). Accidents and decision making under uncertainty: A comparison of four models. Organizational Behavior and Human Decision Processes, 74(2):118 – 144.

Barron, G. and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. Journal of Behavioral Decision Making, 16(3):215–233.

Birnbaum, M. H. and McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. Organizational Behavior and Human Decision Processes, 67(1):91 – 110.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). Inference from Simulations and Monitoring Convergence, chapter 6. CRC press.

Camerer, C. F. and Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. Journal of Risk and Uncertainty, 8(2):167–196.

Camilleri, A. and Newell, B. (2009). The role of representation in experience-based choice. Judgment and Decision Making, 4(9):518–529.

Daunizeau, J., Adam, V., and Rigoux, L. (2014). Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS computational biology, 10(1):e1003441–e1003441. 24465198[pmid].

Fox, C. R. and Hadar, L. (2006). "decisions from experience" = sampling error + prospect theory: Reconsidering hertwig, barron, weber & erev (2004). Judgment and Decision Making, 1(2):159–161.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. Statistics and Computing, 24(6):997–1016.

Glöckner, A., Hilbig, B., Henninger, F., and Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. Journal of Experimental Psychology: General, 145:486–508.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. Psychological Review, 94(2):236–254.

Gonzalez, R. and Wu, G. (1996). Curvature of the probability weighting function. Management Science, 42(12):1676–1690.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. Cognitive Psychology, 38(1):129 – 166.

Hau, R., Pleskac, T. J., and Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. Journal of Behavioral Decision Making, 23(1):48–68.

Hau, R., Pleskac, T. J., Kiefer, J., and Hertwig, R. (2008). The description–experience gap in risky choice: the role of sample size and experienced probabilities. Journal of Behavioral Decision Making, 21(5):493–518.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. Psychological Science, 15(8):534–539. PMID: 15270998.

Hertwig, R. and Erev, I. (2009). The description–experience gap in risky choice. Trends in Cognitive Sciences, 13(12):517 – 523.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2):263–291.

Kellen, D., Pachur, T., and Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? Cognition, 157:126–138.

Lee, M. D. and Wagenmakers, E.-J. (2013). Bayesian cognitive modeling: A practical course. Bayesian cognitive modeling: A practical course. Cambridge University Press, New York, NY, US.

Nilsson, H., Rieskamp, J., and Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. Journal of Mathematical Psychology, 55(1):84 – 93. Special Issue on Hierarchical Bayesian Models.

Peters, O., Adamou, A., Kirstein, M., and Berman, Y. (2020). What are we weighting for? A mechanistic model for probability weighting. arXiv e-prints, page arXiv:2005.00056.

Prelec, D. (1998). The probability weighting function. Econometrica, 66(3):497–527.

Preston, M. G. and Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. The American Journal of Psychology, 61:183–193.

Rakow, T., Demes, K. A., and Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. Organizational Behavior and Human Decision Processes, 106(2):168 – 179.

Steiner, J. and Stewart, C. (2016). Perceiving prospects properly. The American Economic Review, 106(7):1601–1631.

Stewart, N., Chater, N., and Brown, G. D. A. (2006). Decision by sampling.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. Journal of Risk and Uncertainty, 32(2):101–130.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. Psychological Review, 102(2):269–283.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5:297–323.

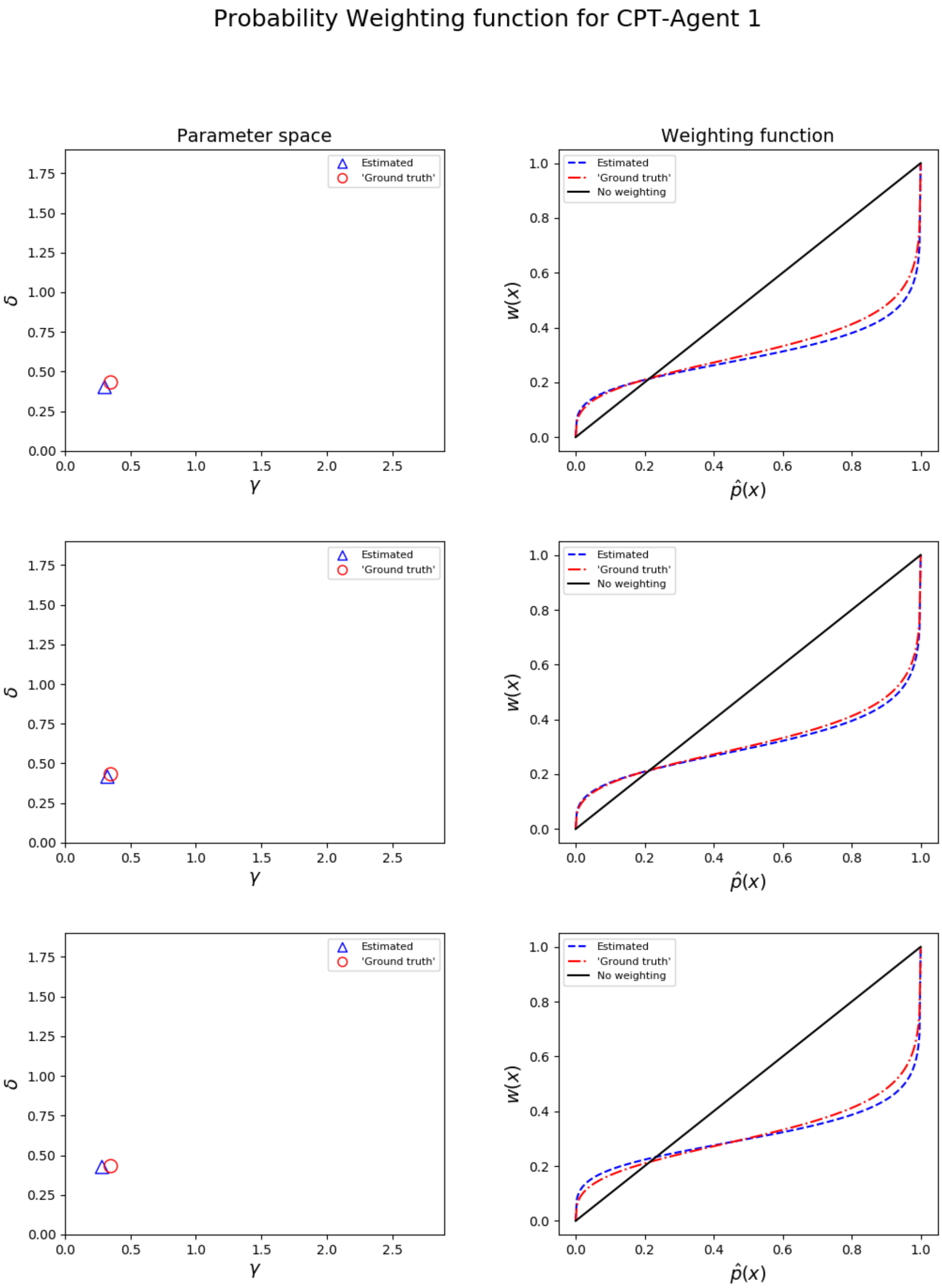
Ungemach, C., Chater, N., and Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? Psychological Science, 20(4):473–479.

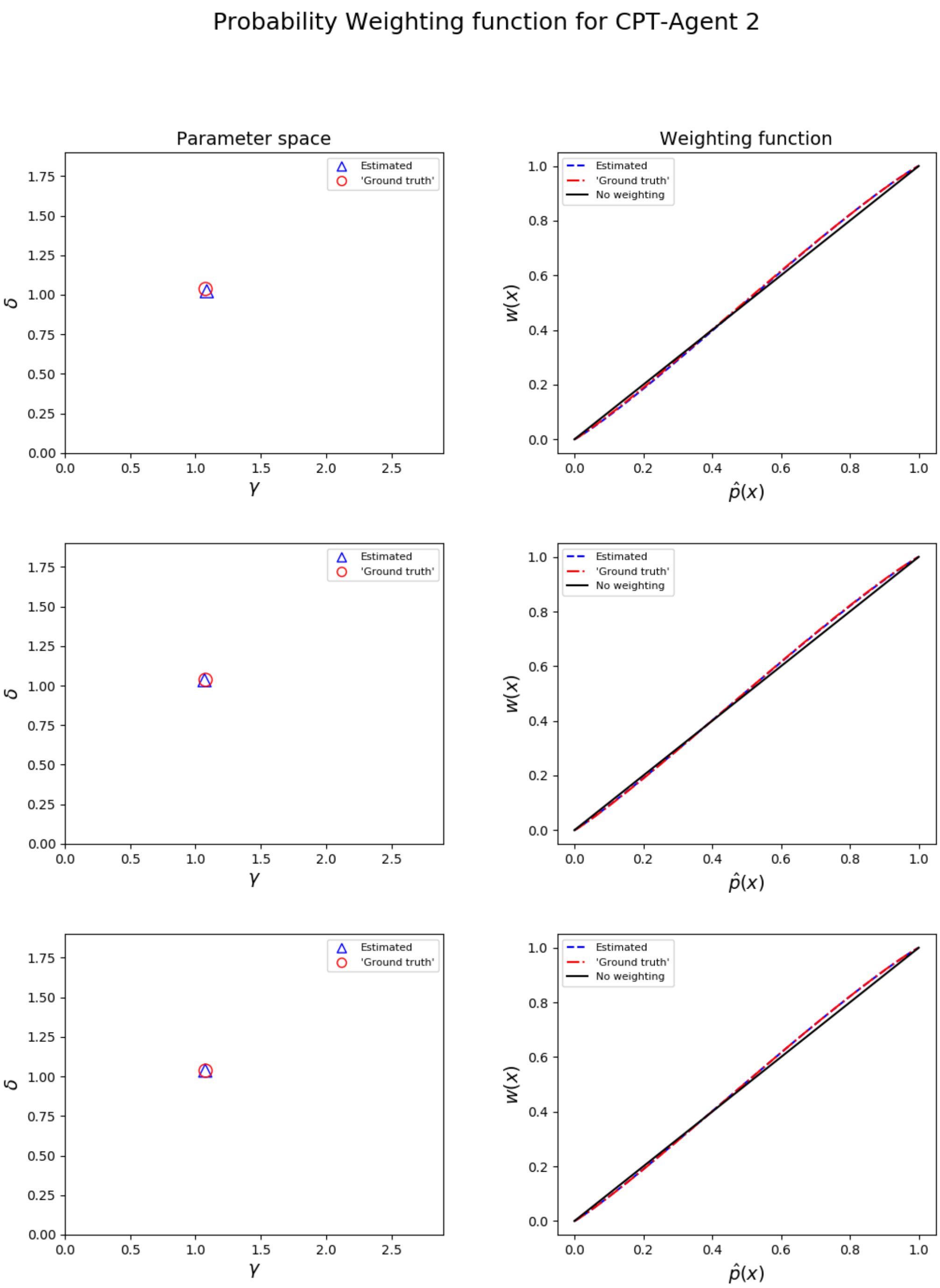
Weber, E., Shafir, S., and Blais, A.-R. (2004). Predicting risk-sensitivity in humans and lower animals: Risk as variance or coefficient of variation. In Psychological review.

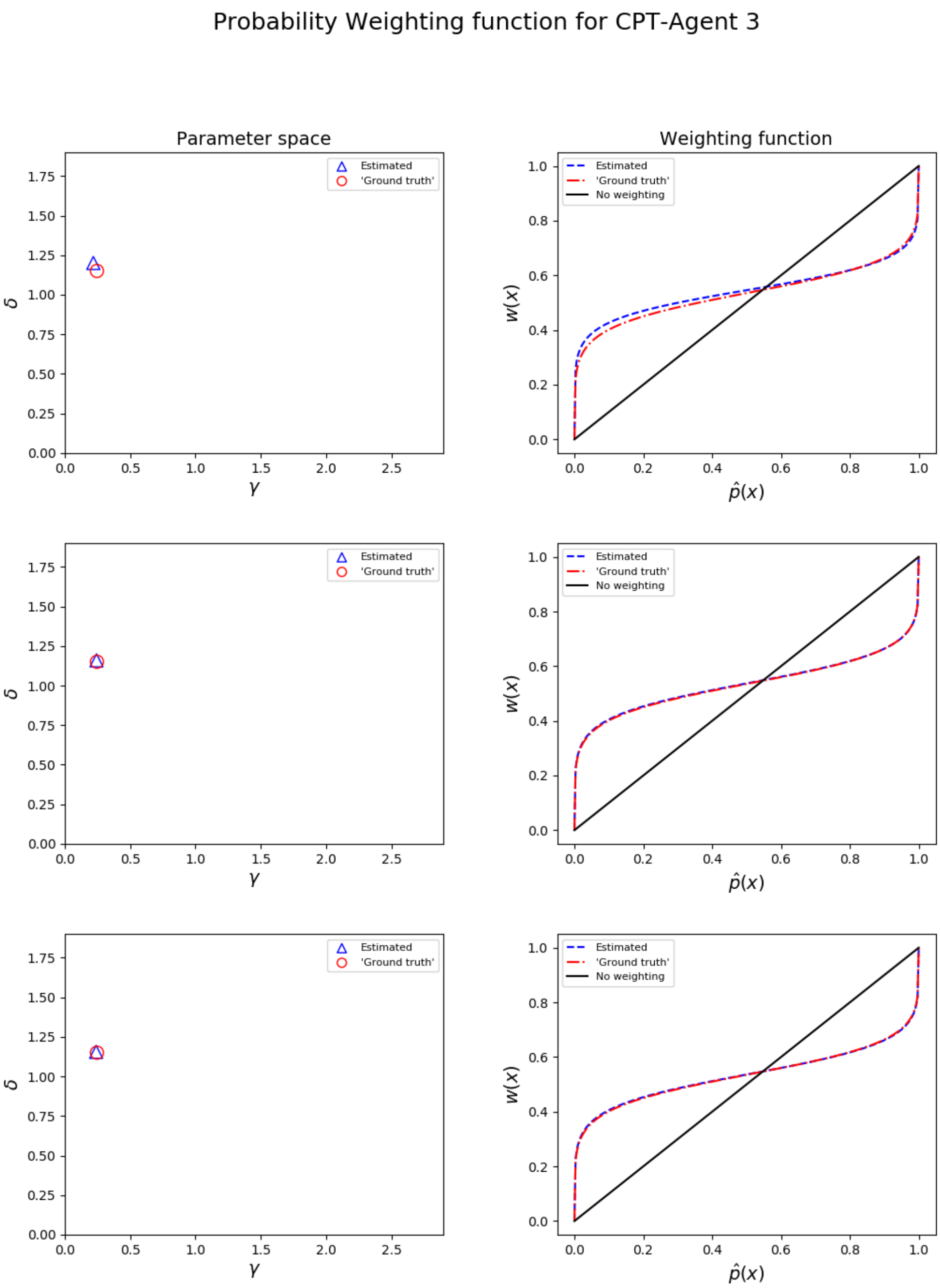
Wulff, D. U., Mergenthaler-Canseco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. Psychological Bulletin, 144(2):140–176.

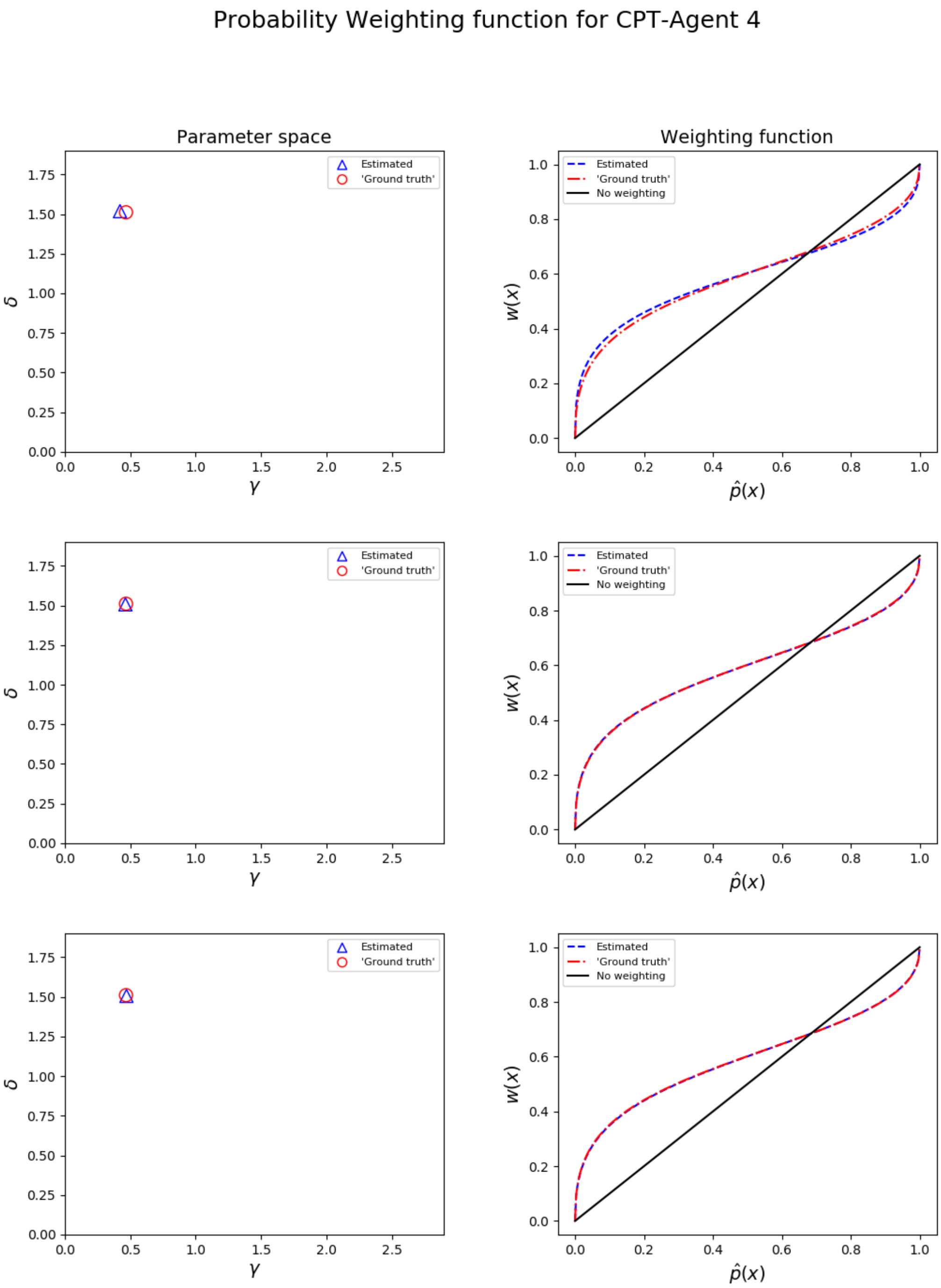
A Individual PW functions for all agents

A.1 CPT-agents

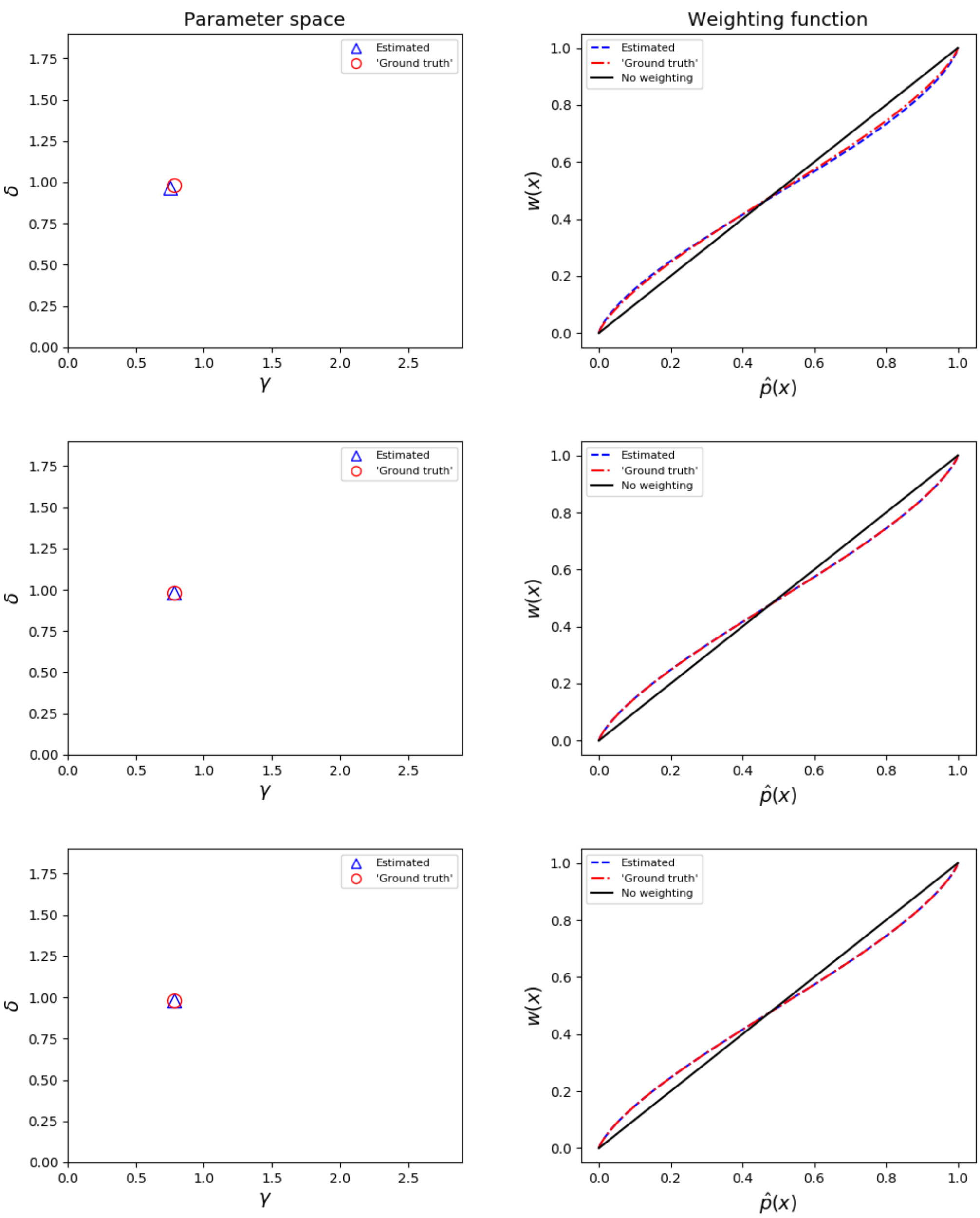


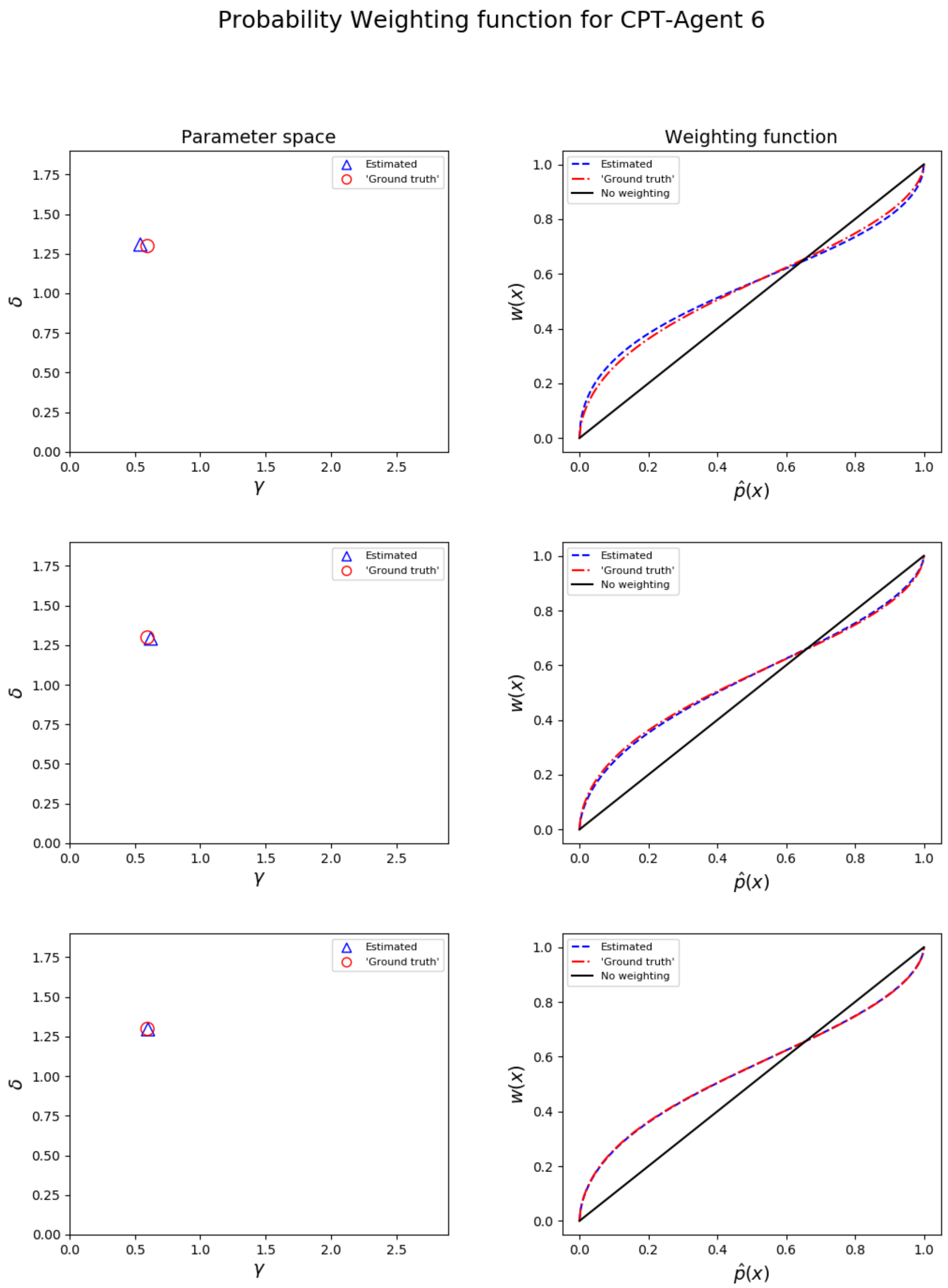

Probability Weighting function for CPT-Agent 2
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8

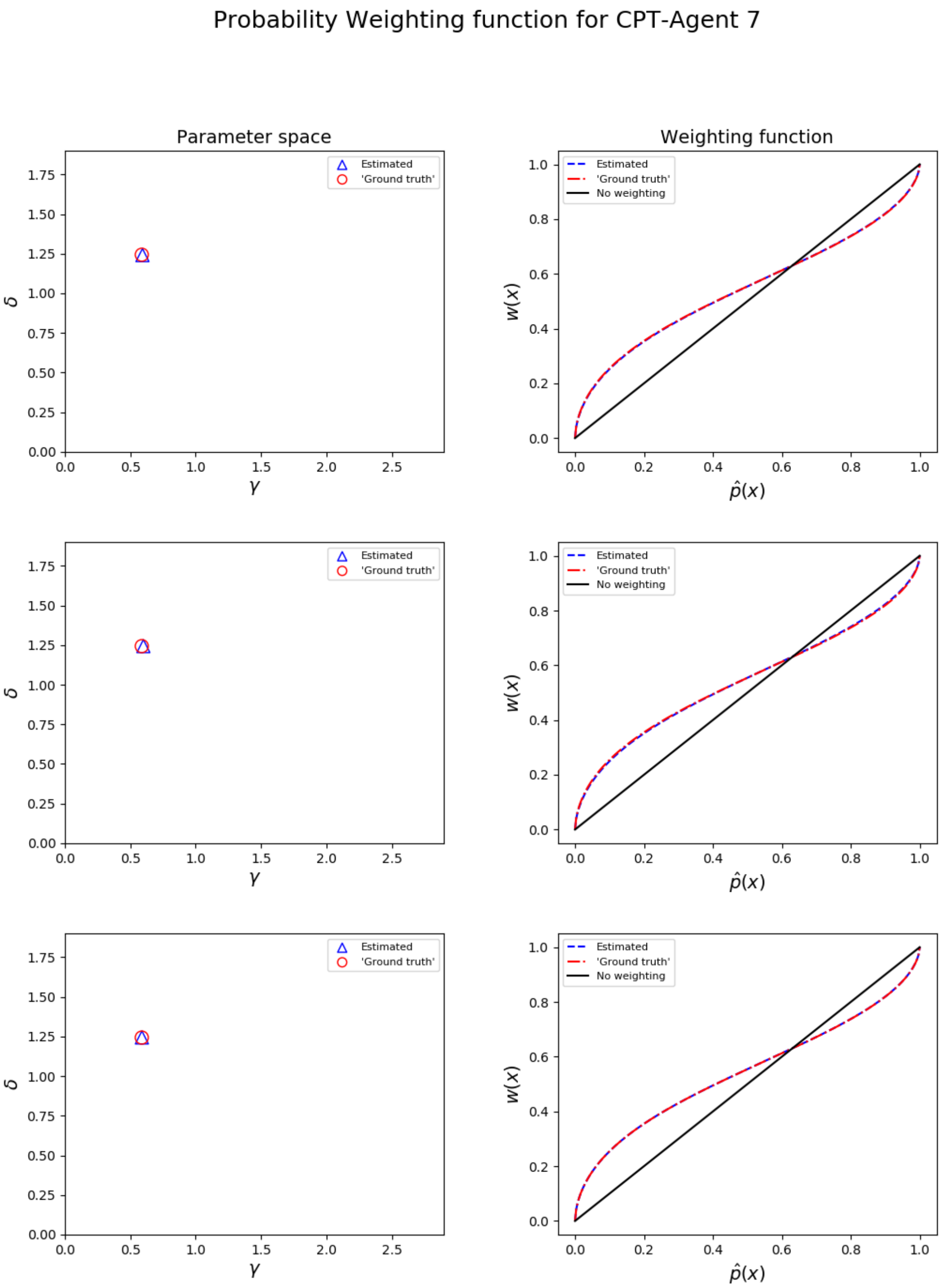

Probability Weighting function for CPT-Agent 3
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)

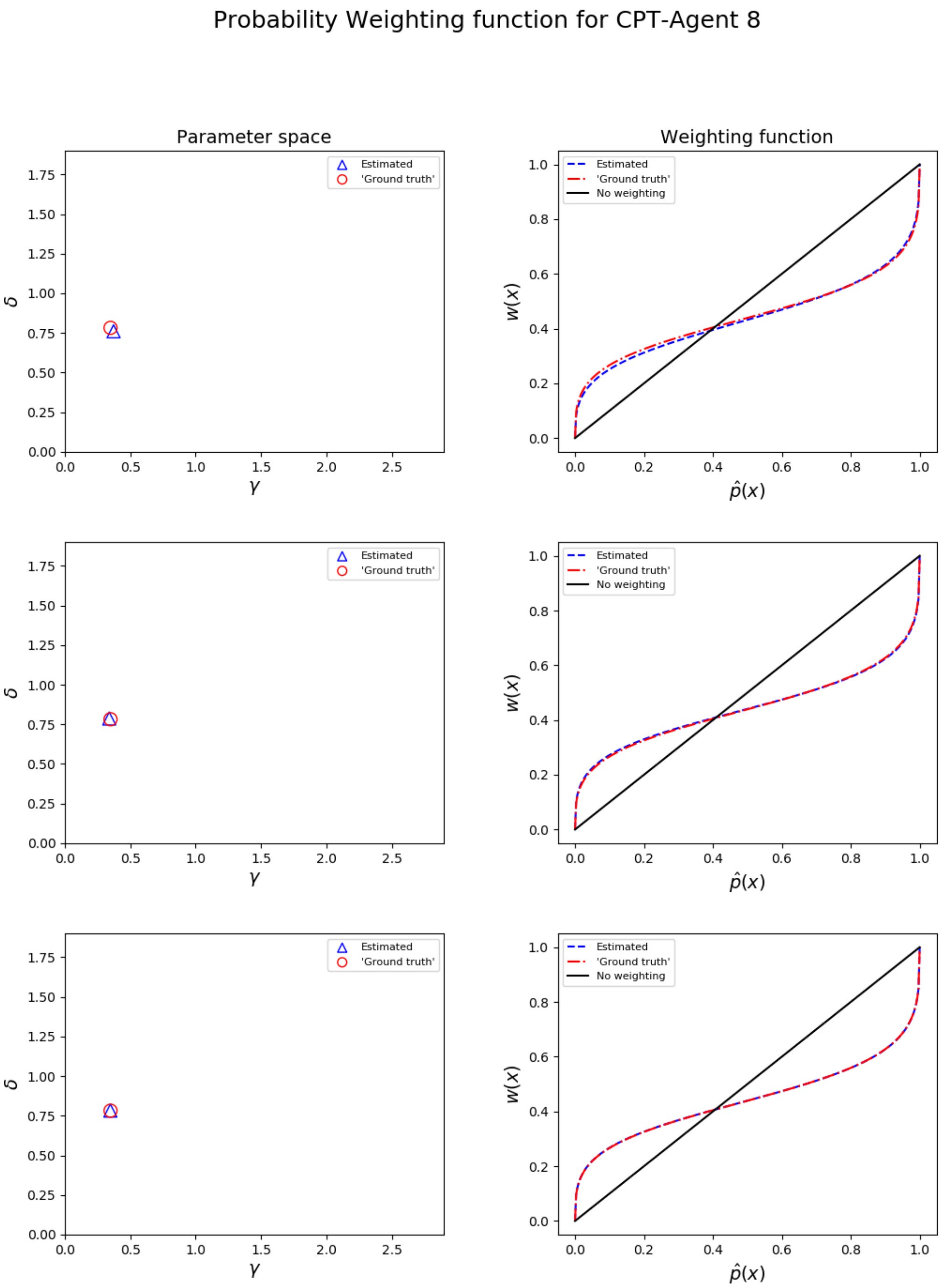

Probability Weighting function for CPT-Agent 4
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
γ
δ
w(x)
p̂(x)
0.0
0.5
1.0
1.5
2.0
2.5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.2
0.4
0.6
0.8

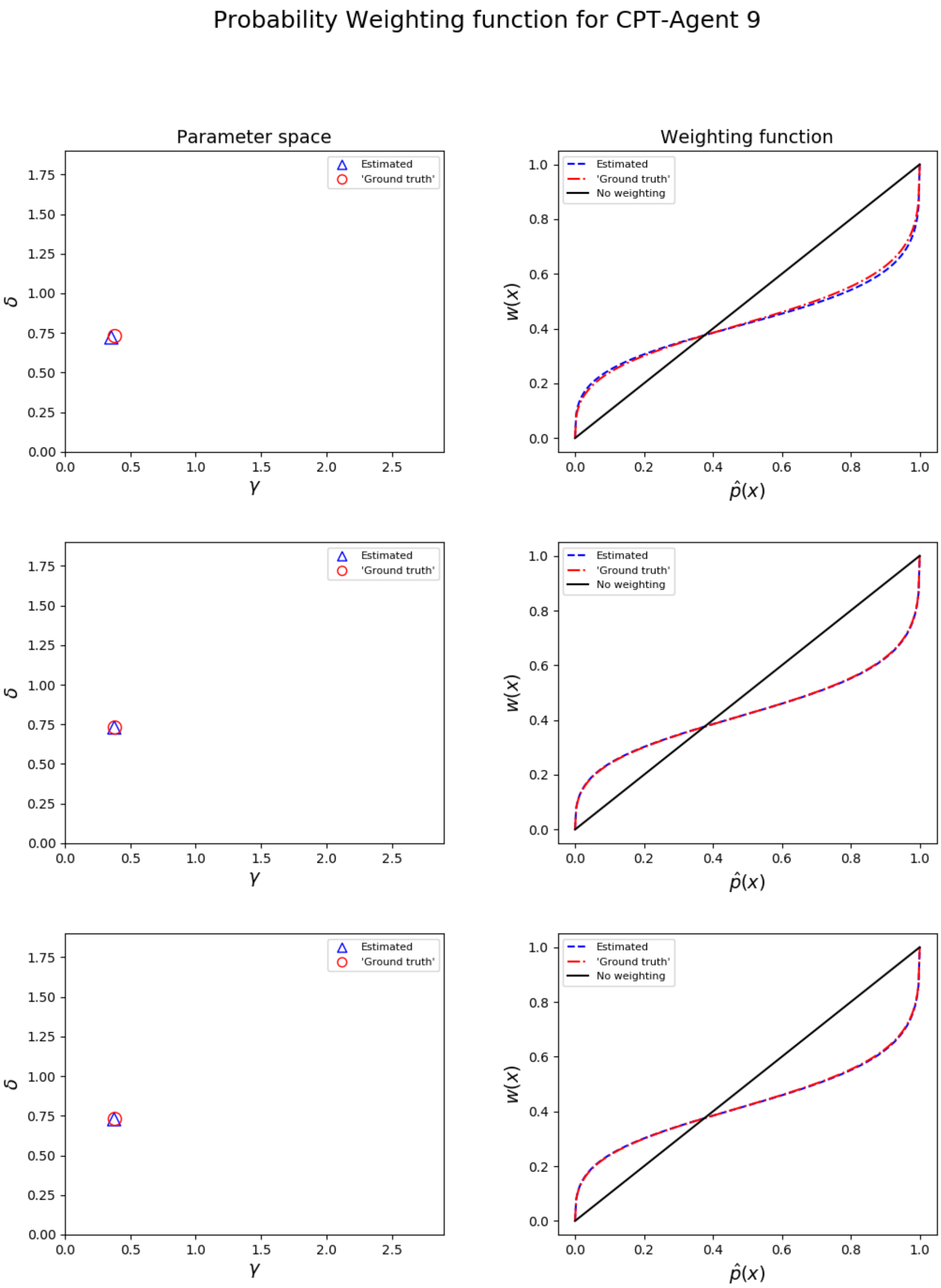
Probability Weighting function for CPT-Agent 5



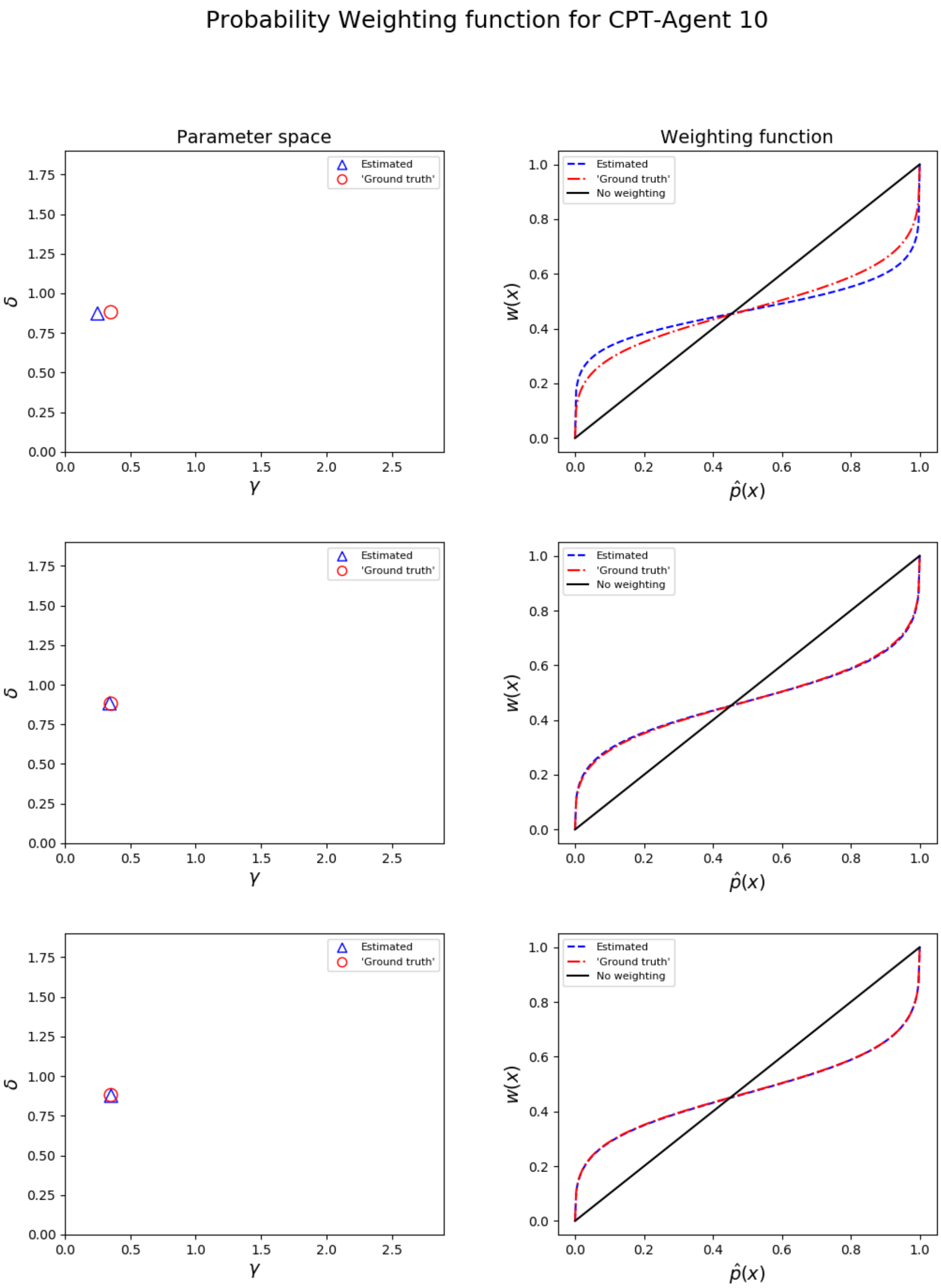

Probability Weighting function for CPT-Agent 6
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8

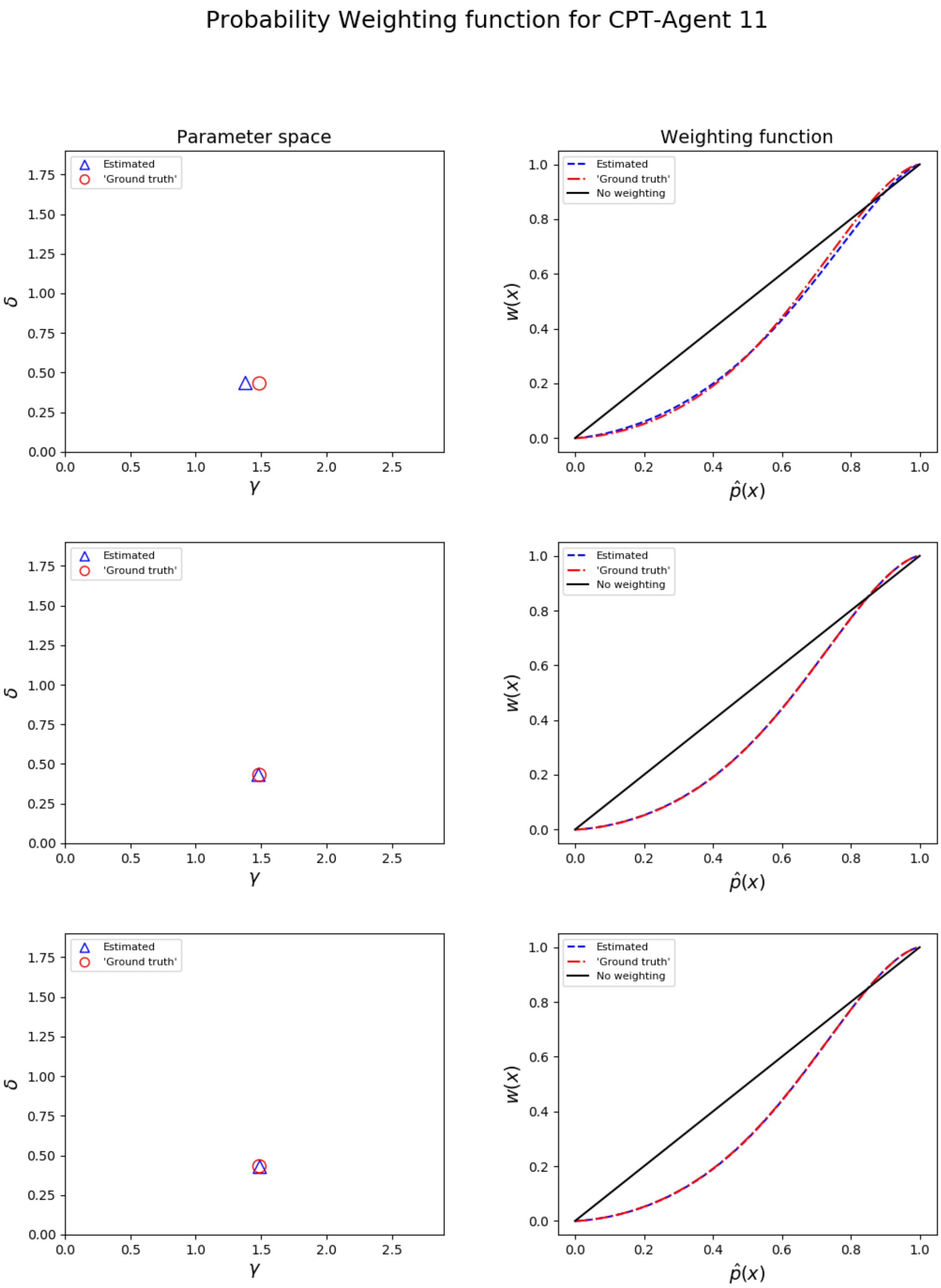

Probability Weighting function for CPT-Agent 7
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)

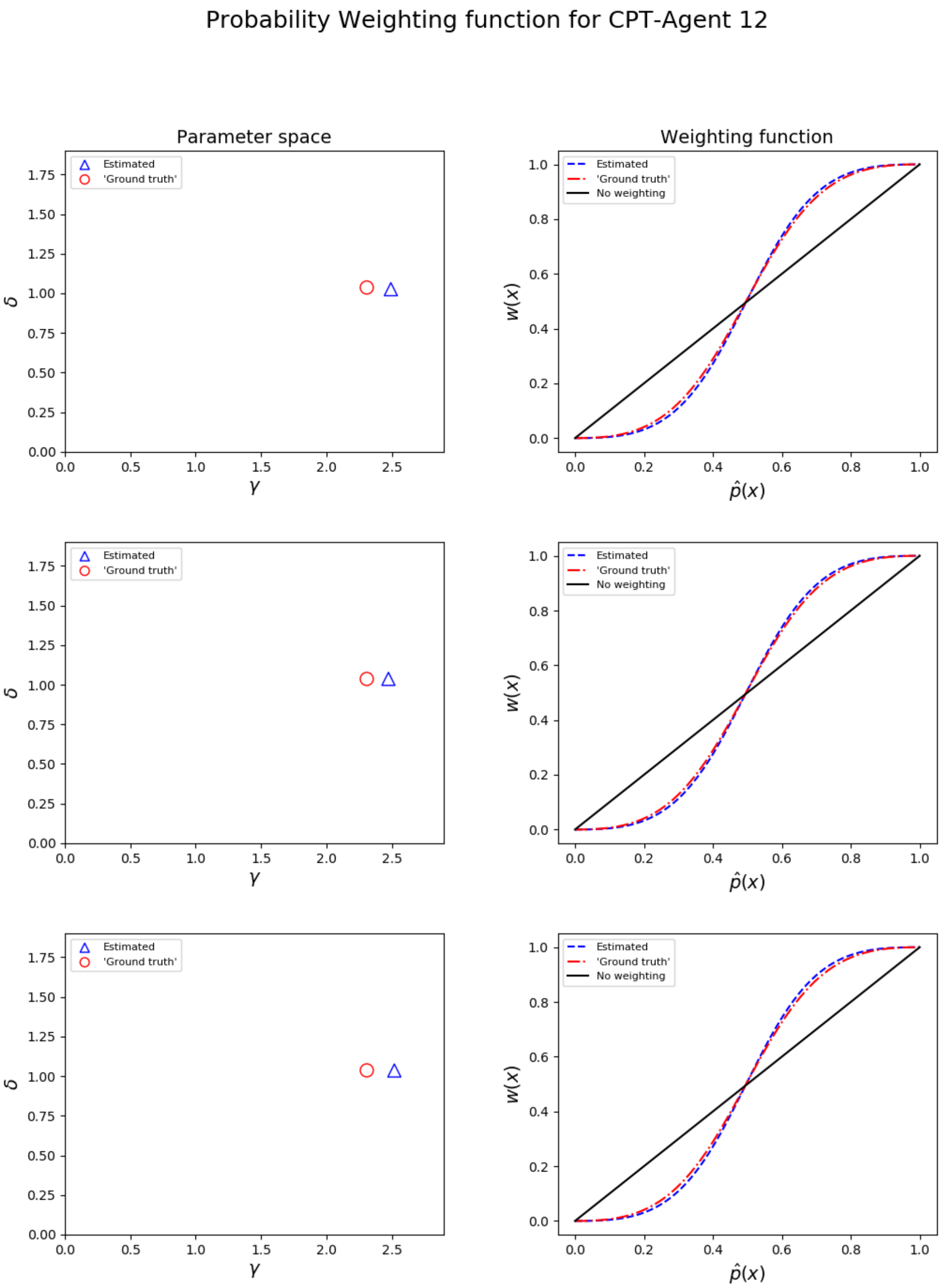

Probability Weighting function for CPT-Agent 8
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)

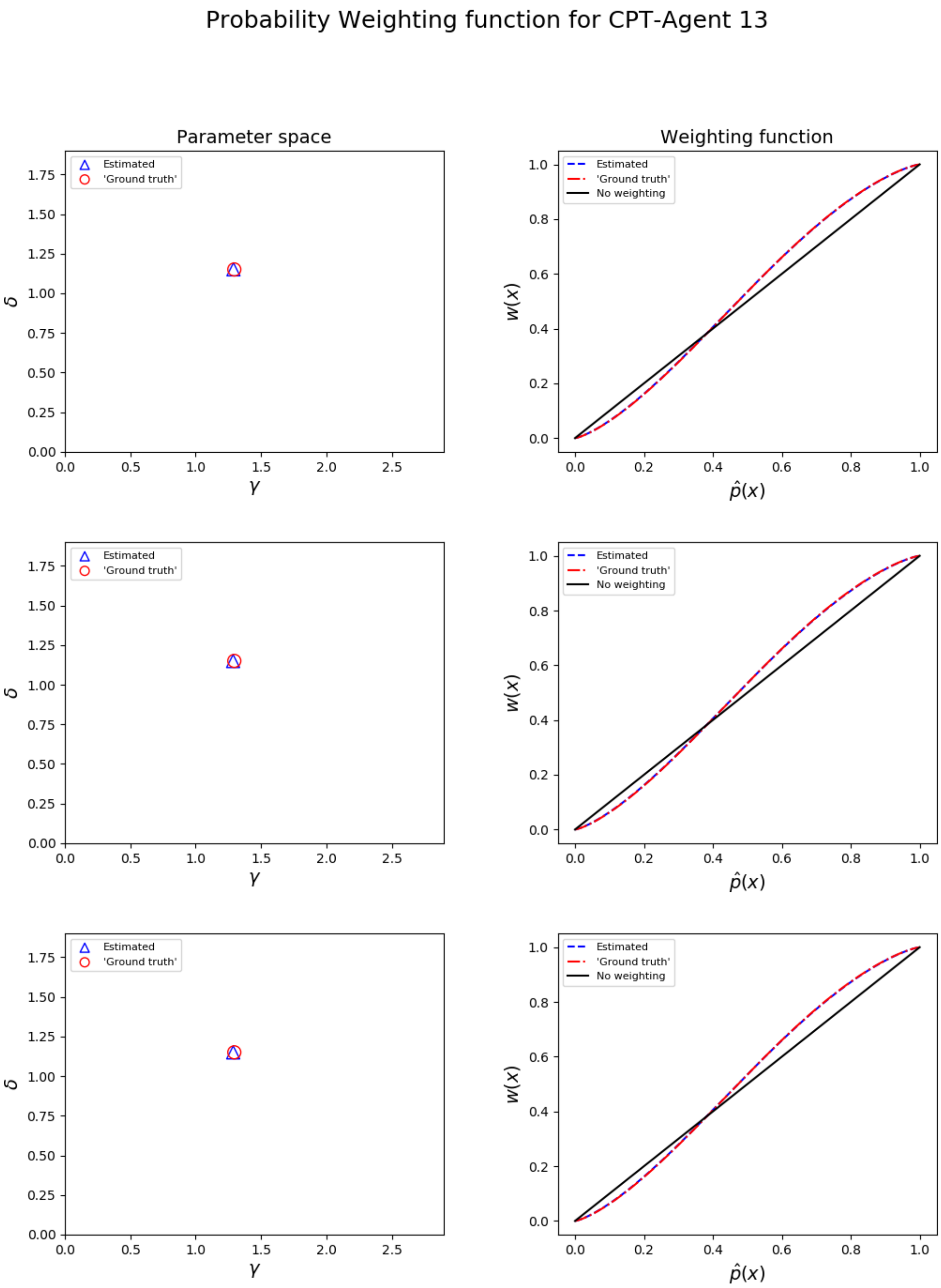


Probability Weighting function for CPT-Agent 9
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)

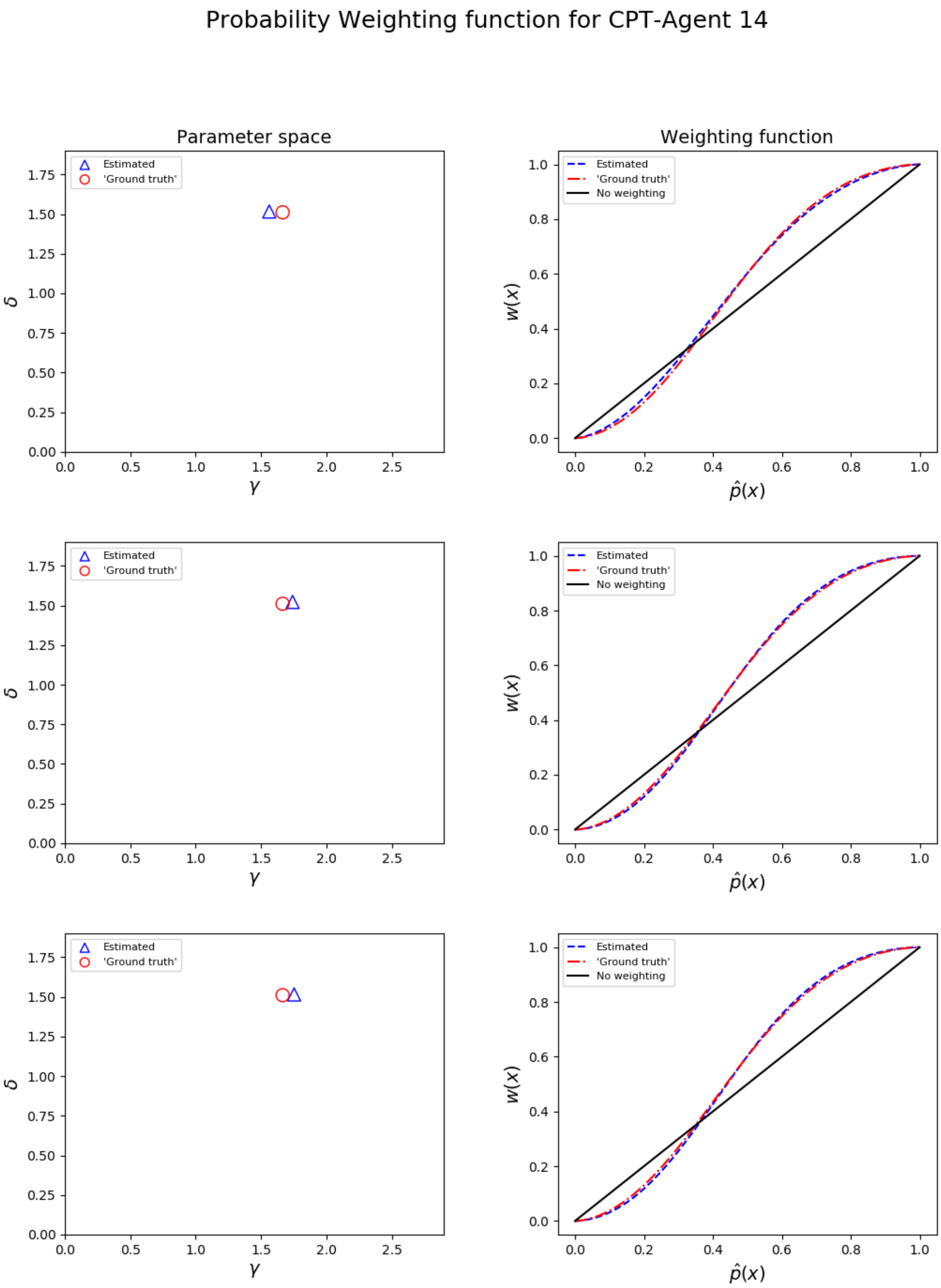

Probability Weighting function for CPT-Agent 10
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)


Probability Weighting function for CPT-Agent 11
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8

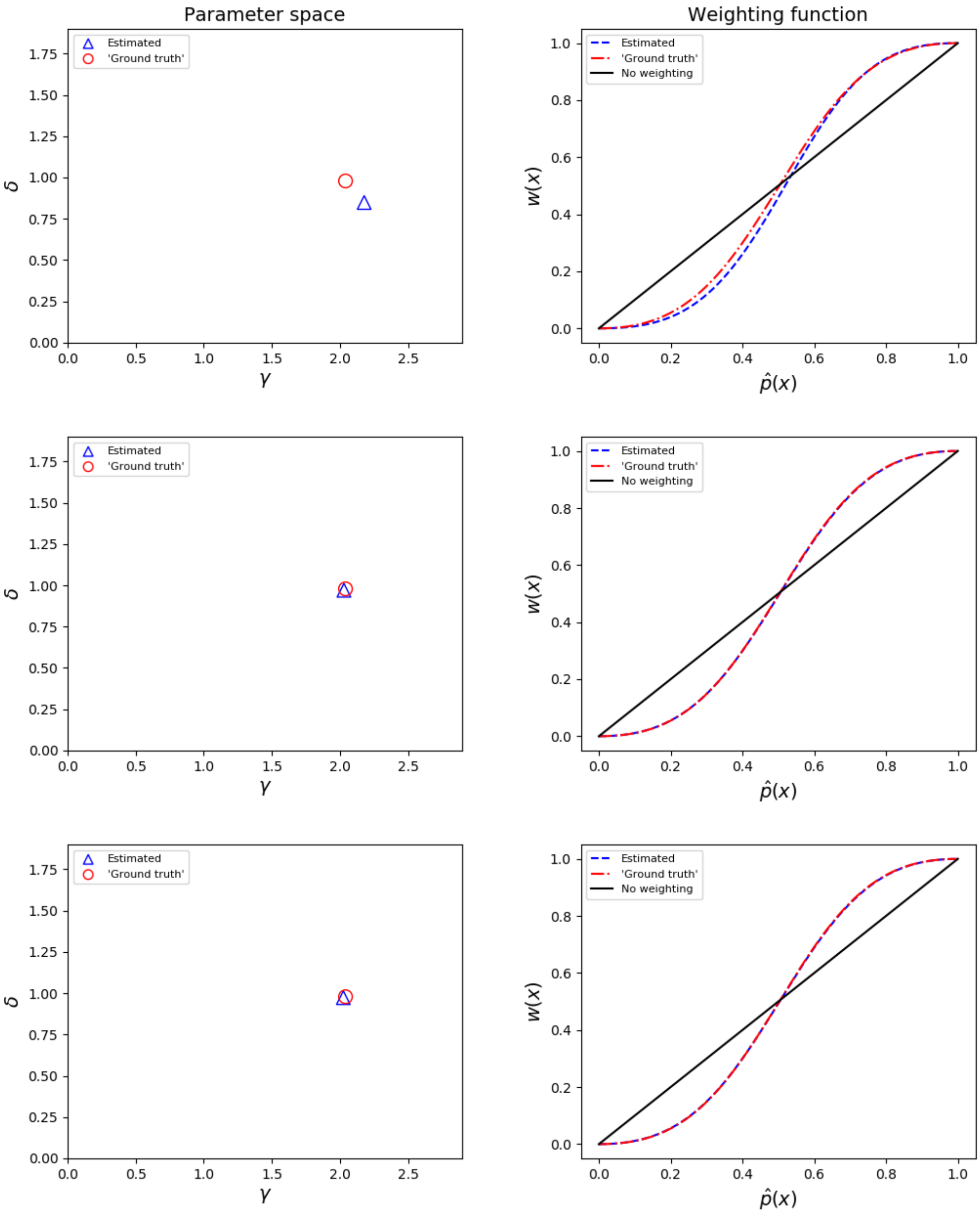

Probability Weighting function for CPT-Agent 12
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
$w(x)$
$\hat{p}(x)$

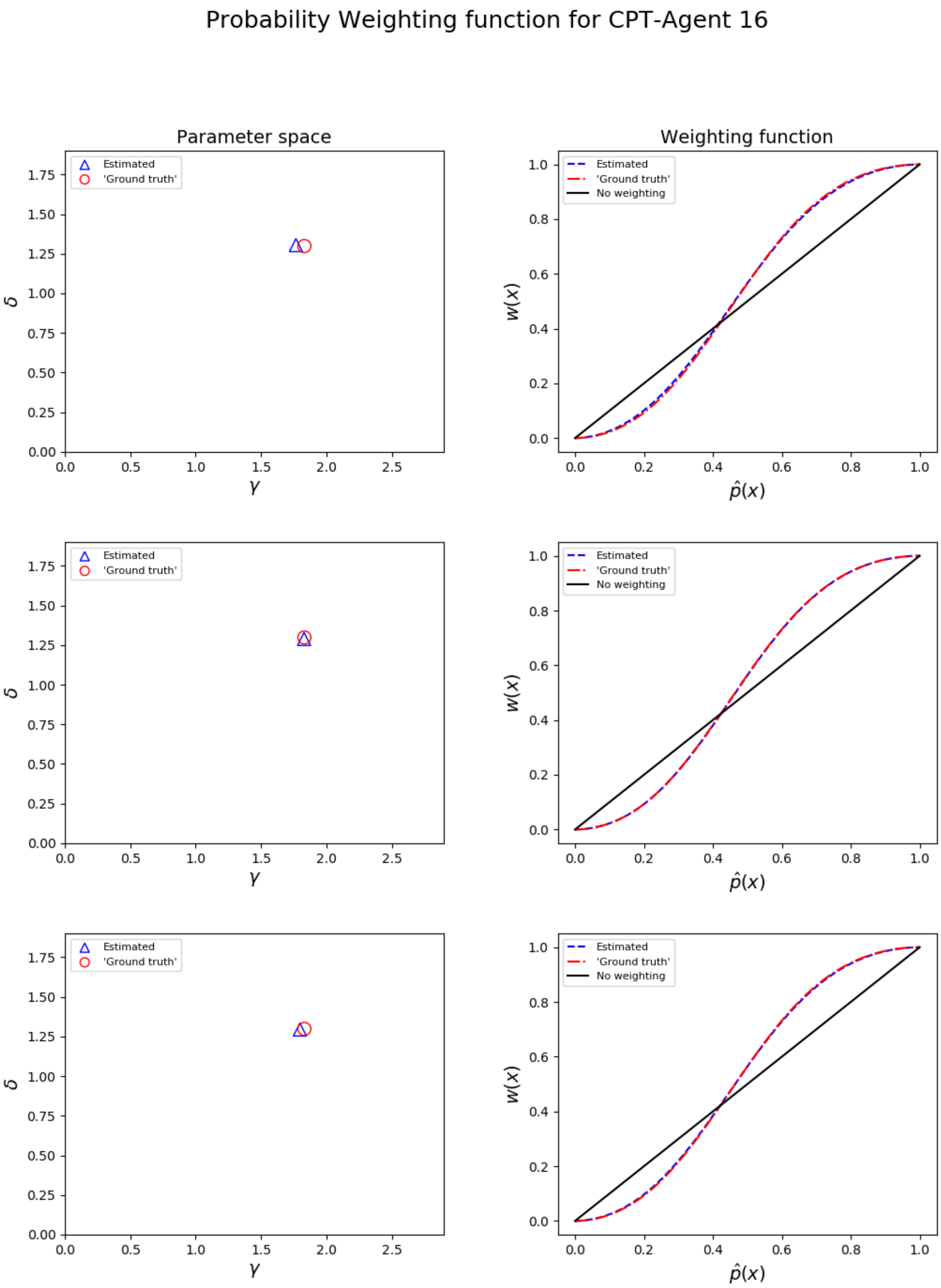


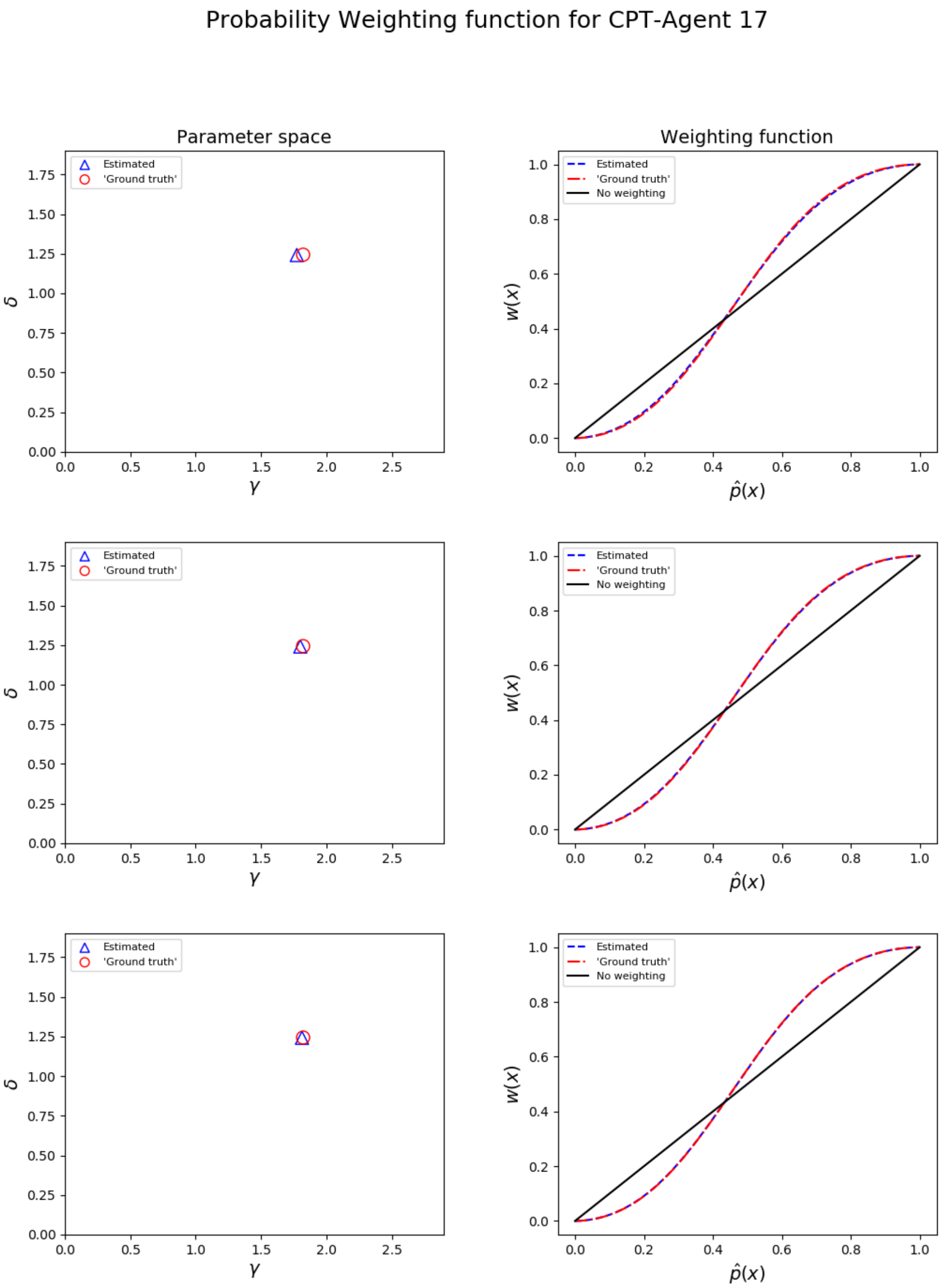
Probability Weighting function for CPT-Agent 13
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
$w(x)$
$\hat{p}(x)$

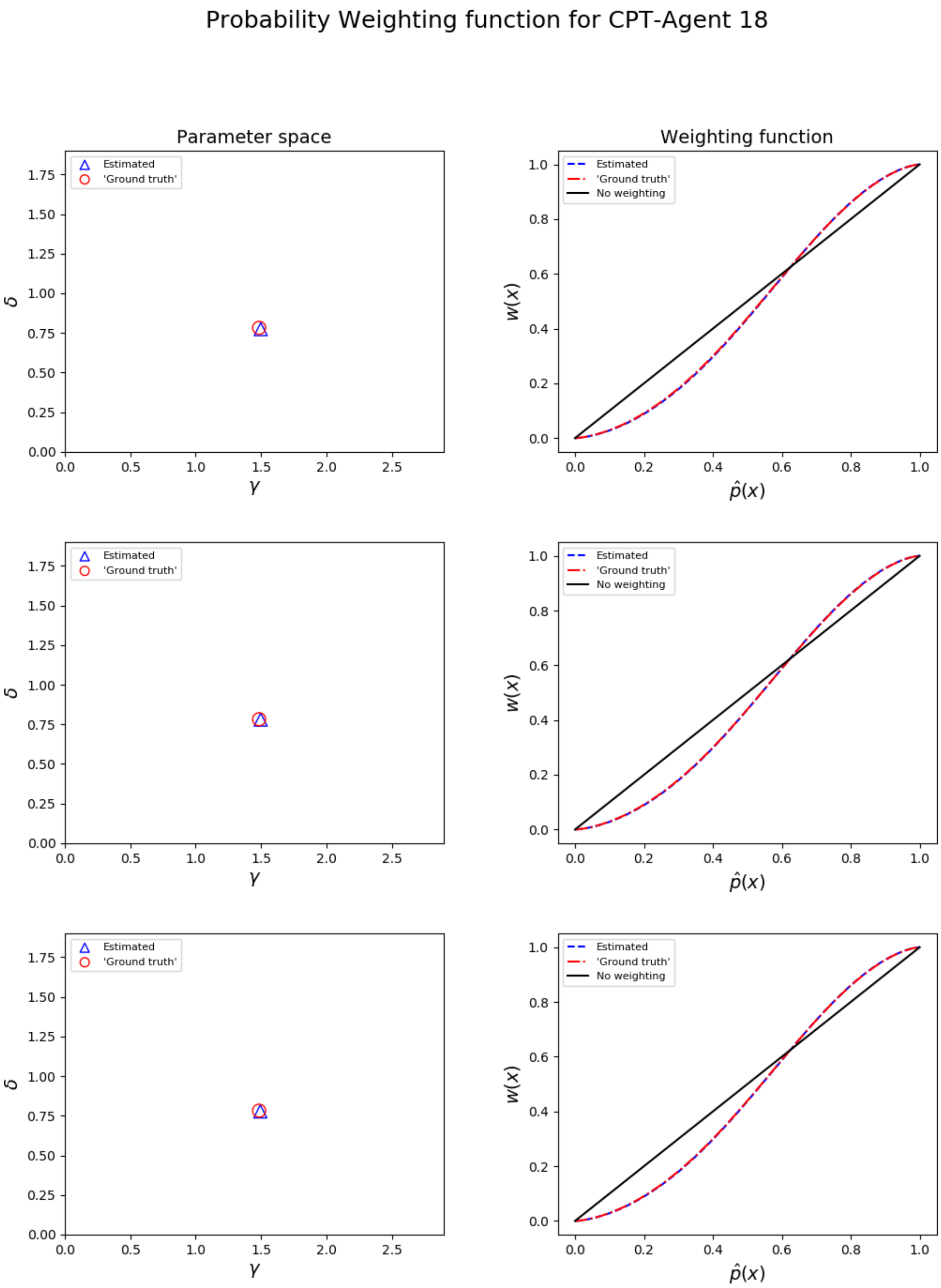

Probability Weighting function for CPT-Agent 14
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)

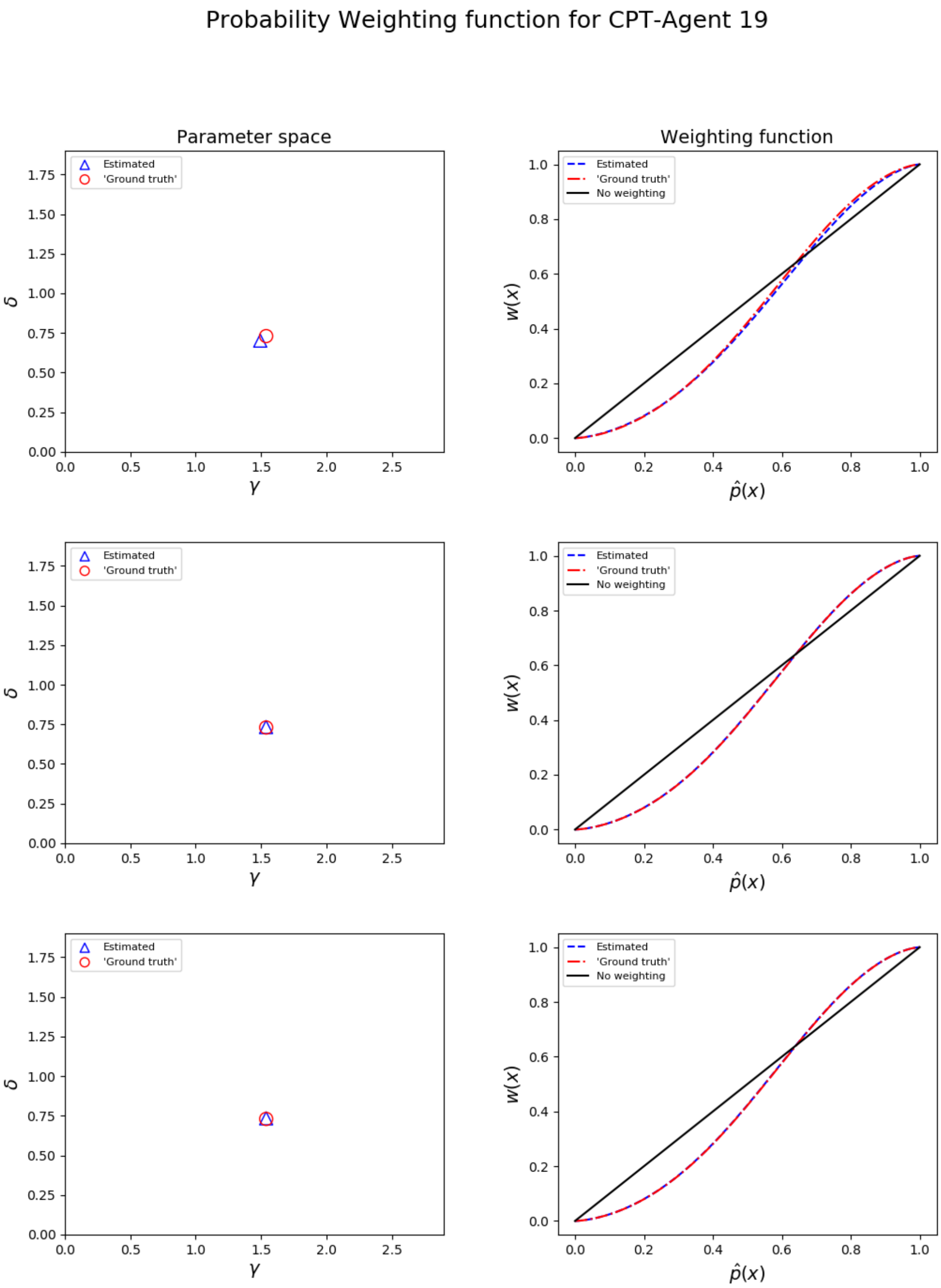
Probability Weighting function for CPT-Agent 15



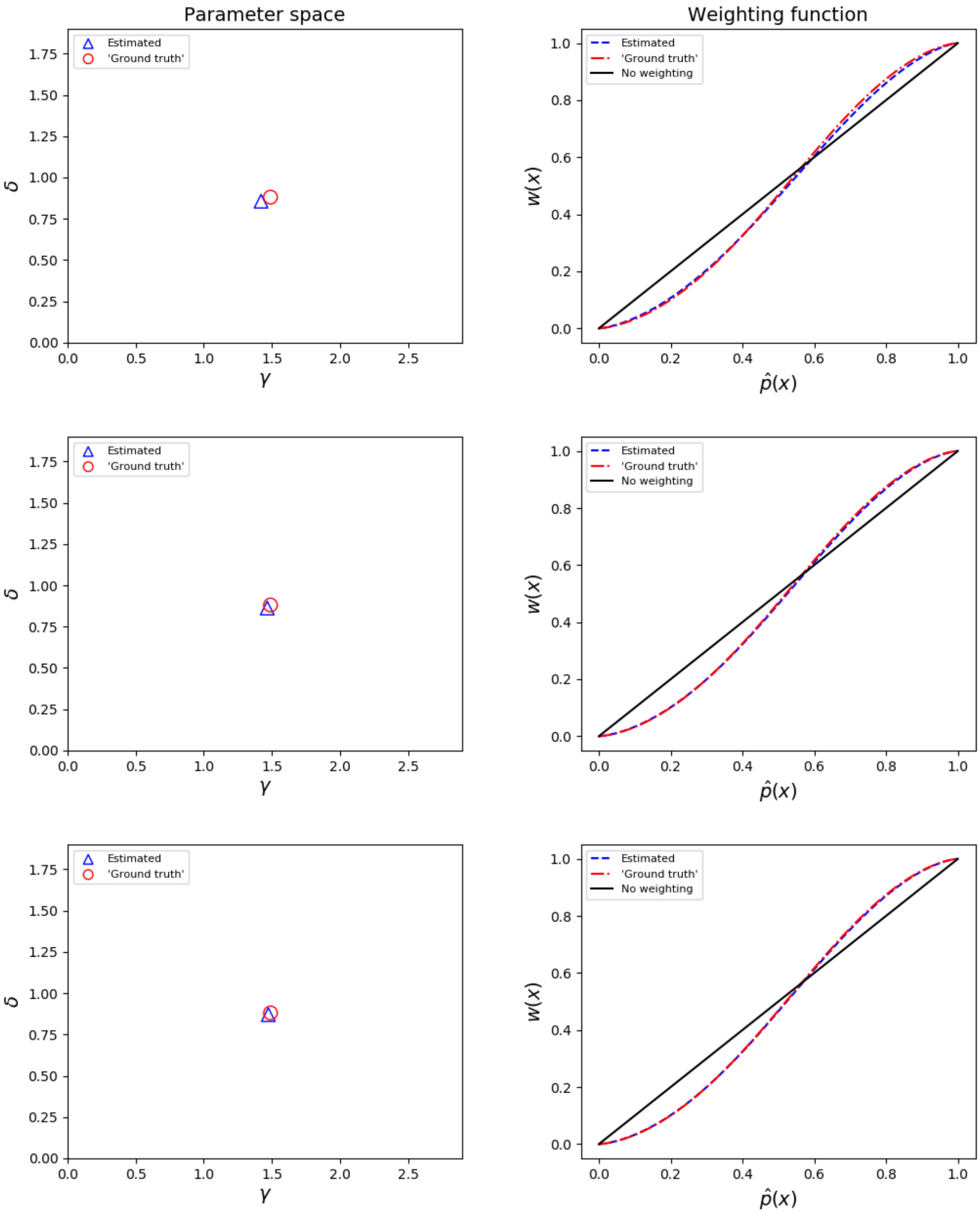

Probability Weighting function for CPT-Agent 16
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
p̂(x)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8


Probability Weighting function for CPT-Agent 17
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
γ
δ
$w(x)$
$\hat{p}(x)$

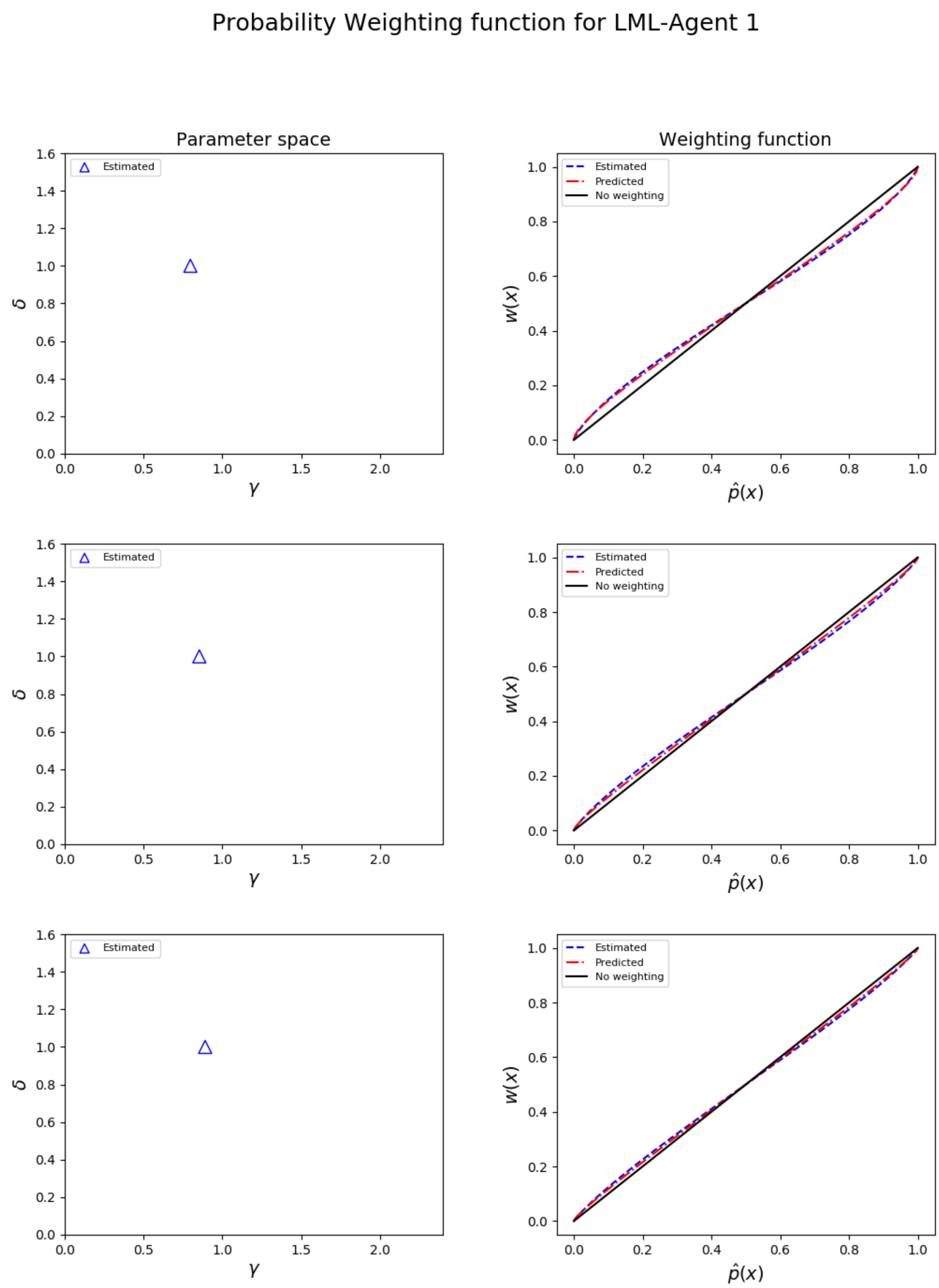

Probability Weighting function for CPT-Agent 18
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
w(x)
$\hat{p}(x)$

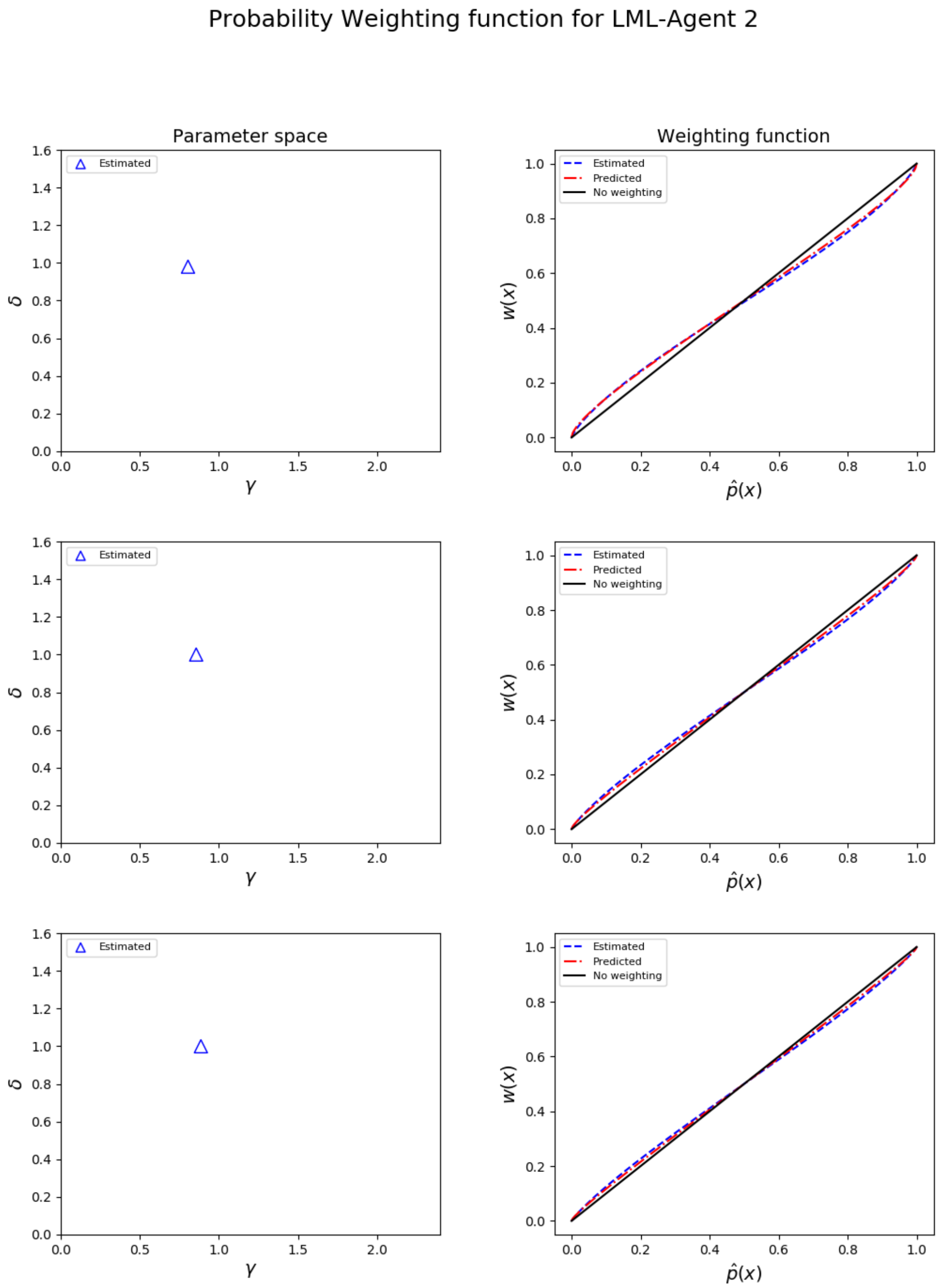

Probability Weighting function for CPT-Agent 19
Parameter space
Weighting function
Estimated
'Ground truth'
No weighting
δ
γ
$w(x)$
$\hat{p}(x)$

Probability Weighting function for CPT-Agent 20

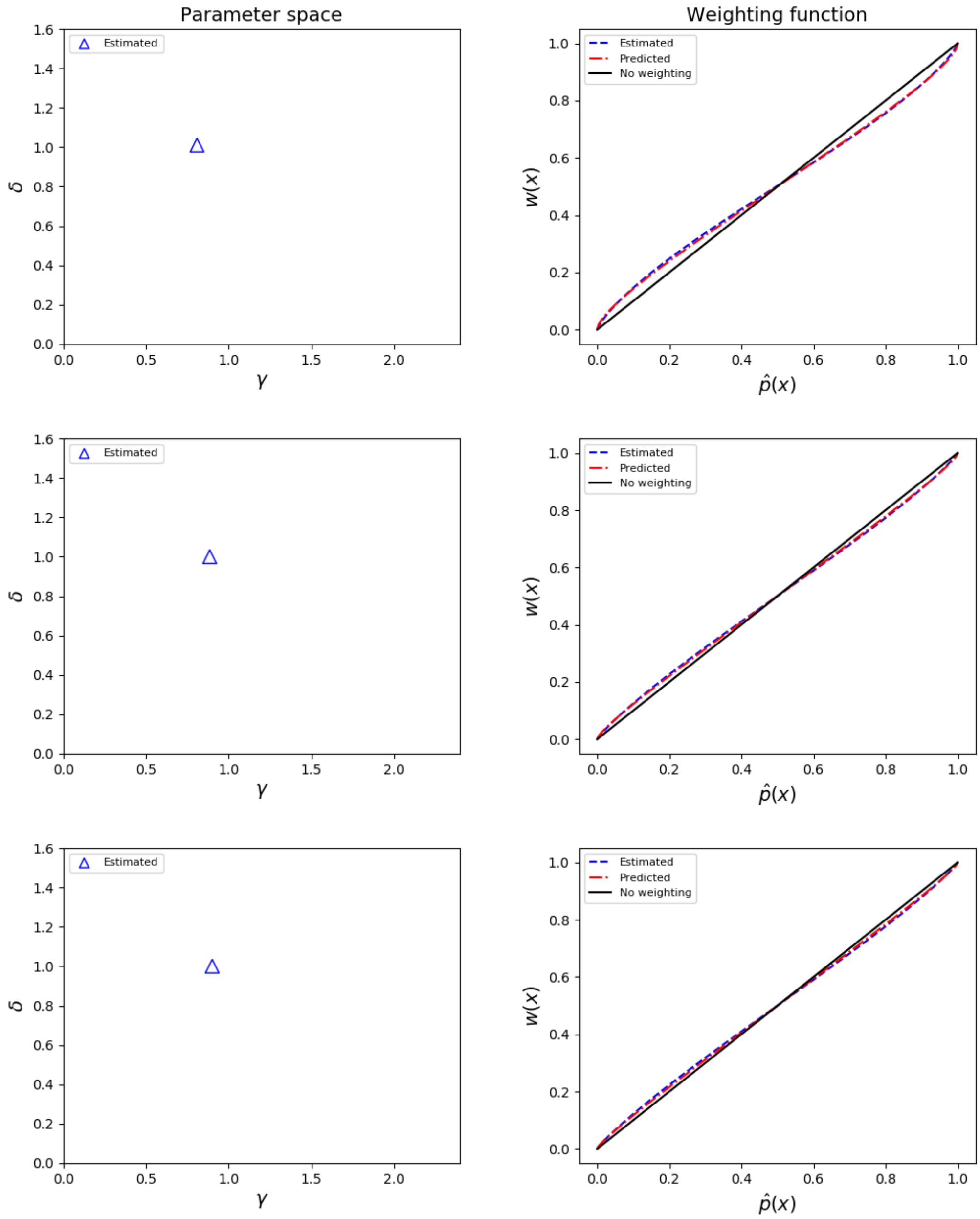


A.2 Mechanistic-agents

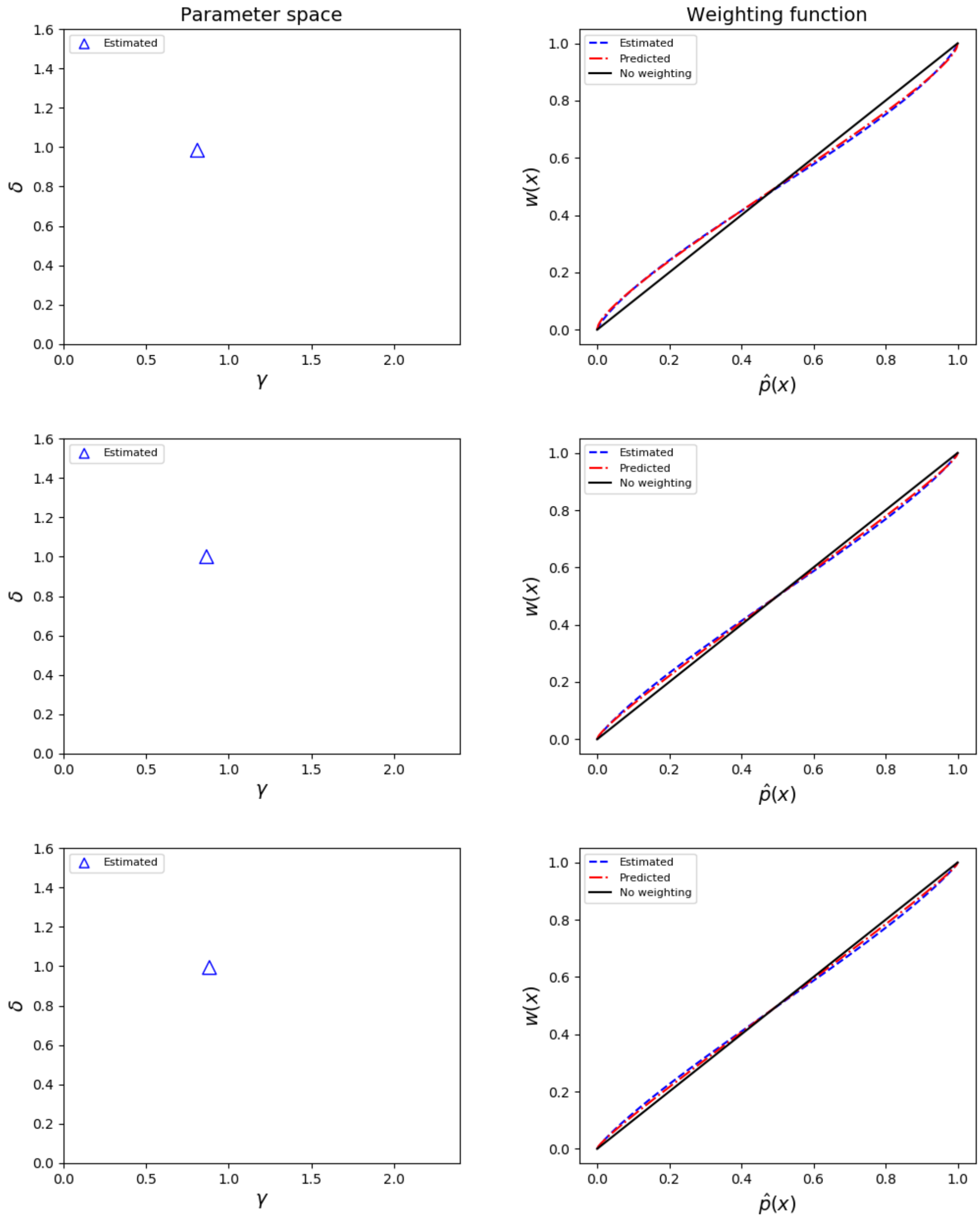



Probability Weighting function for LML-Agent 2
Parameter space
Weighting function
Estimated
Predicted
No weighting
δ
γ
$w(x)$
$\hat{p}(x)$
0.0
0.2
0.4
0.5
0.6
0.8
1.0
1.2
1.4
1.5
1.6
2.0

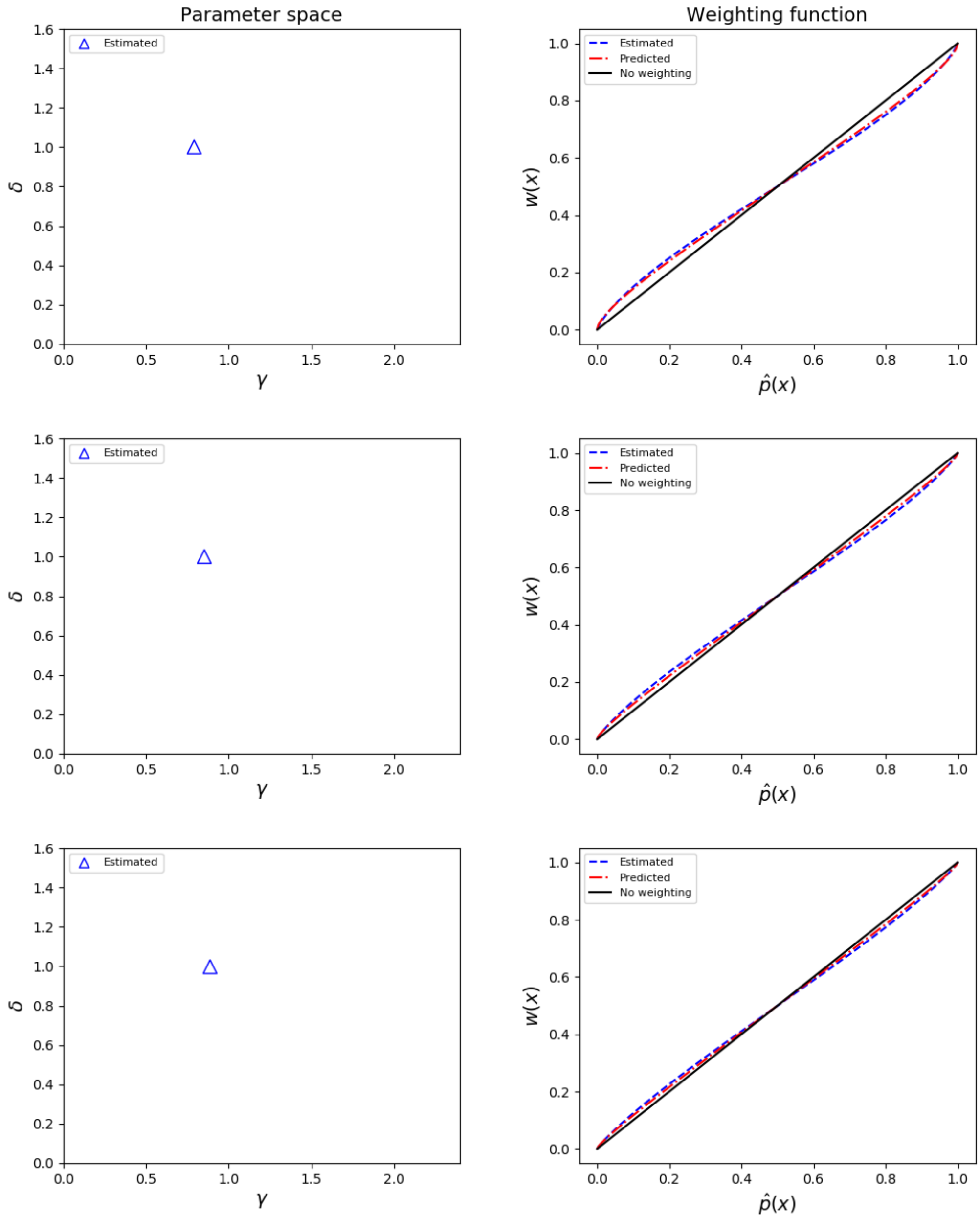
Probability Weighting function for LML-Agent 3

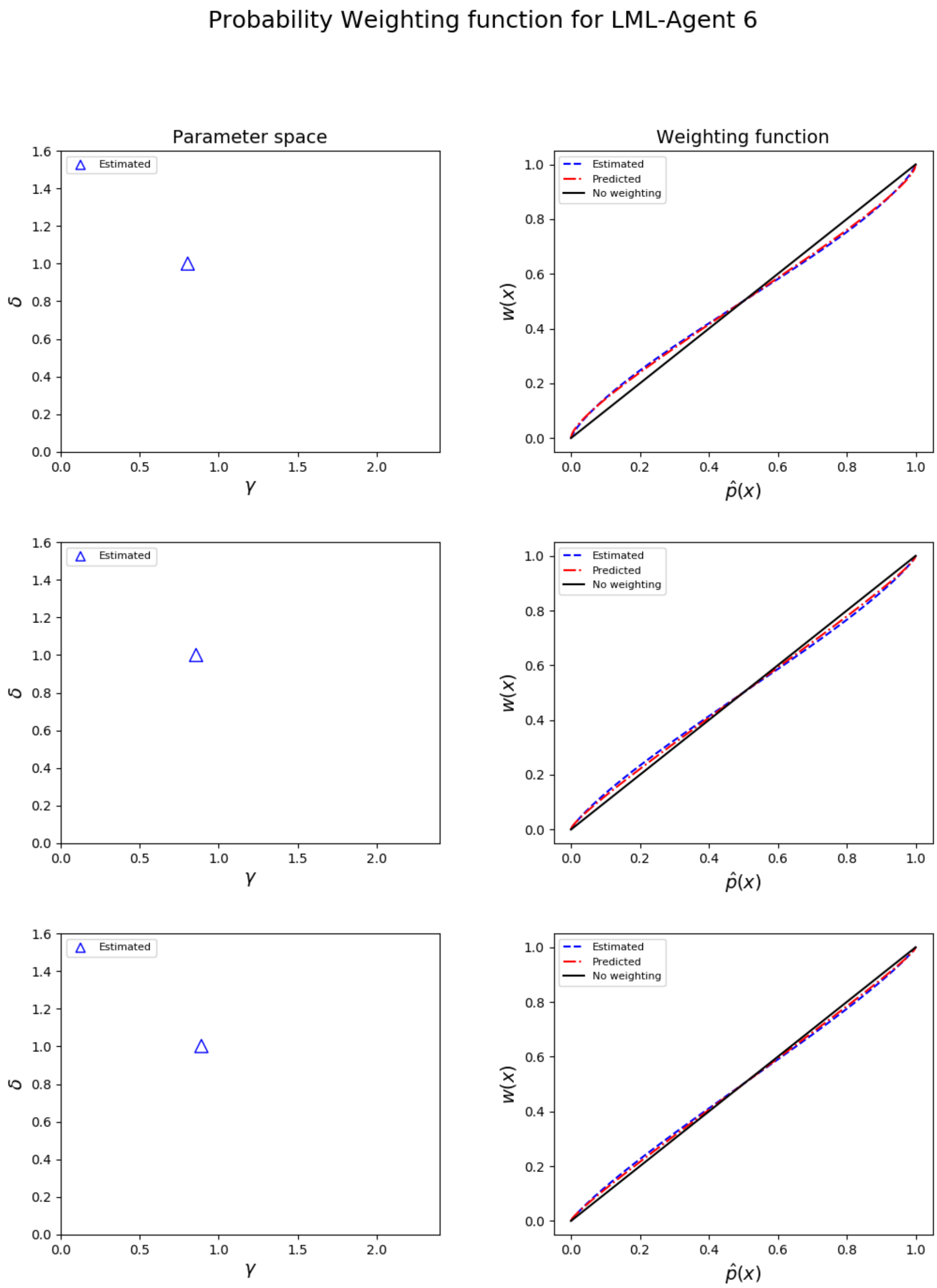


Probability Weighting function for LML-Agent 4

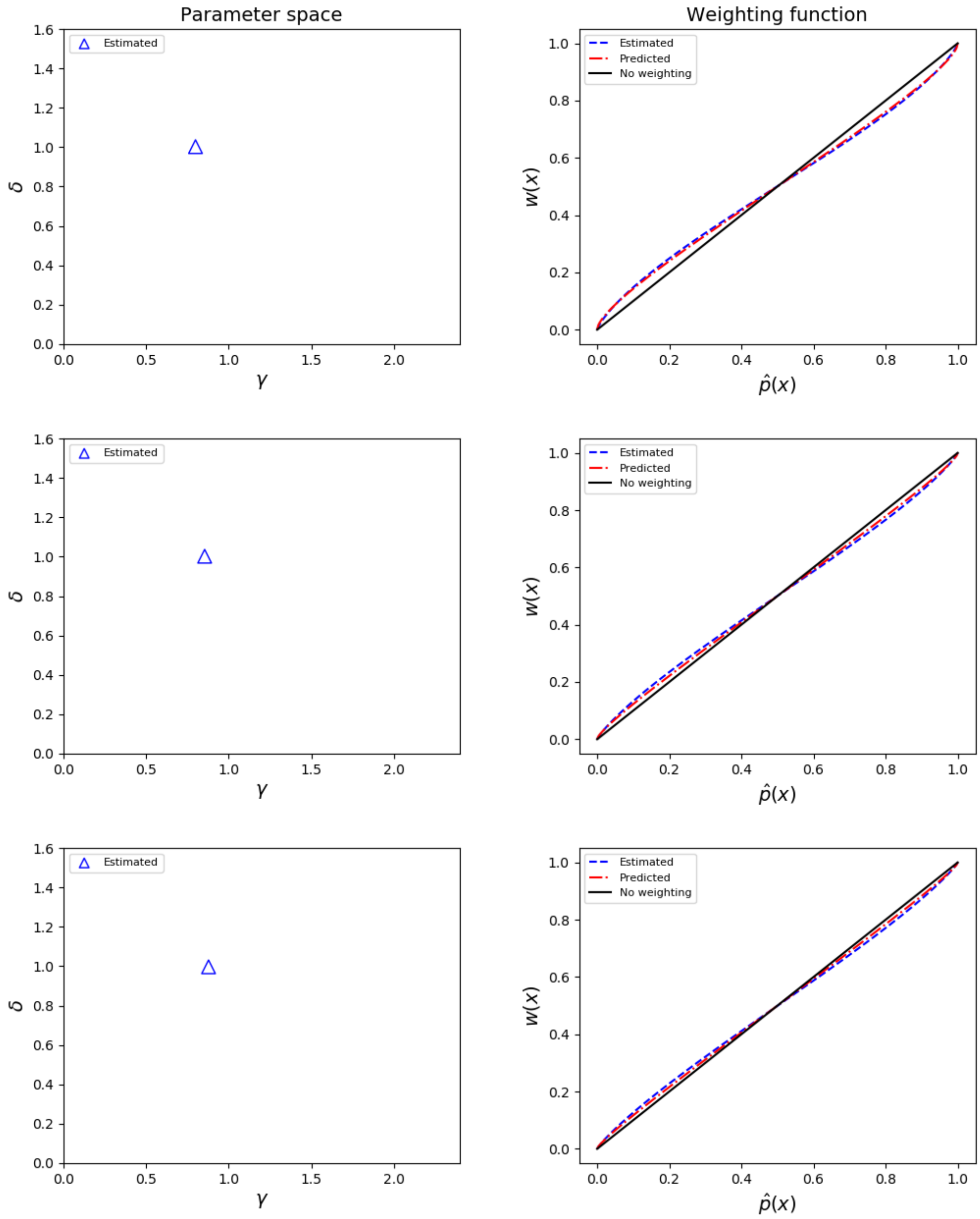


Probability Weighting function for LML-Agent 5

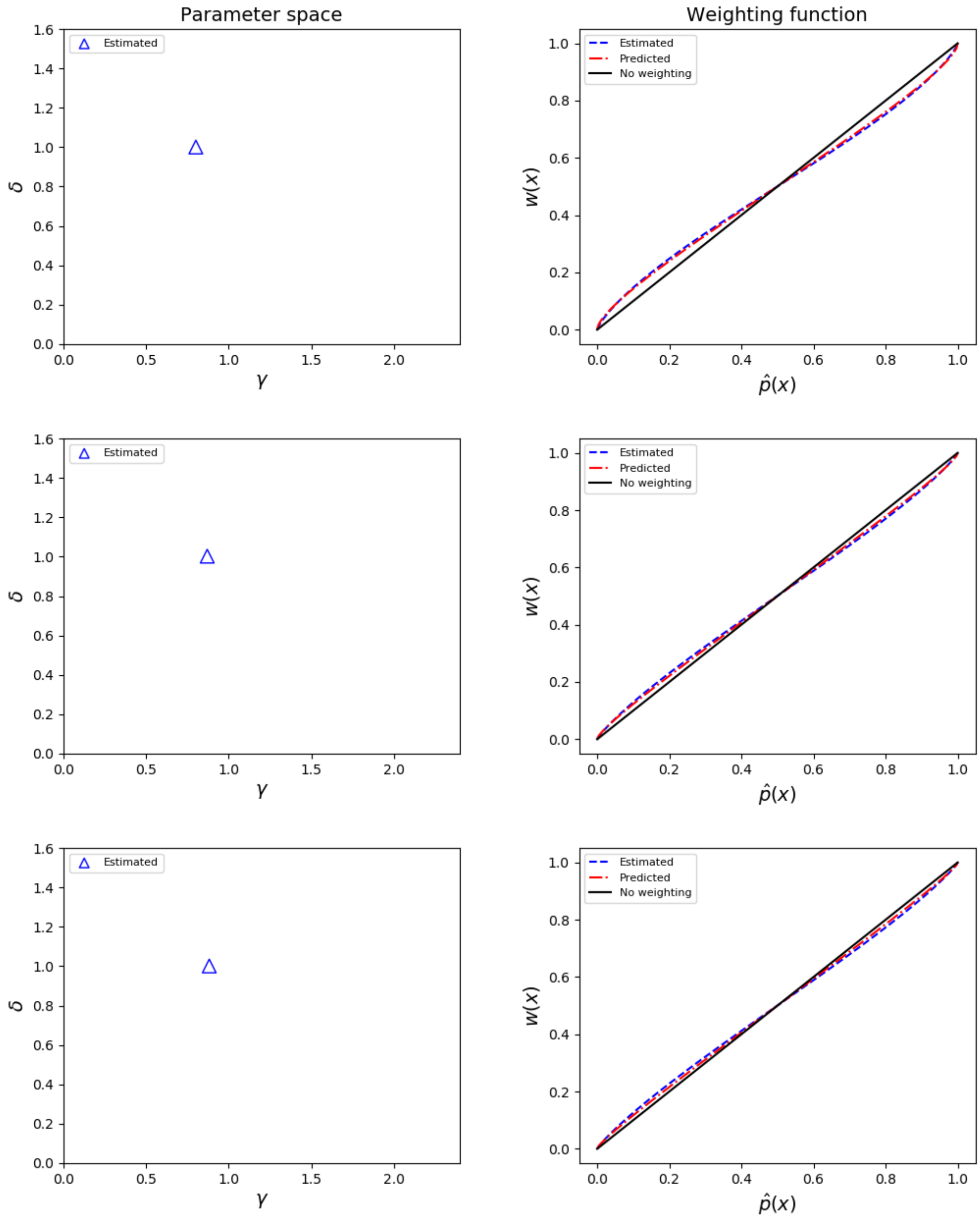



Probability Weighting function for LML-Agent 6
Parameter space
Weighting function
Estimated
Predicted
No weighting
δ
γ
w(x)
p̂(x)

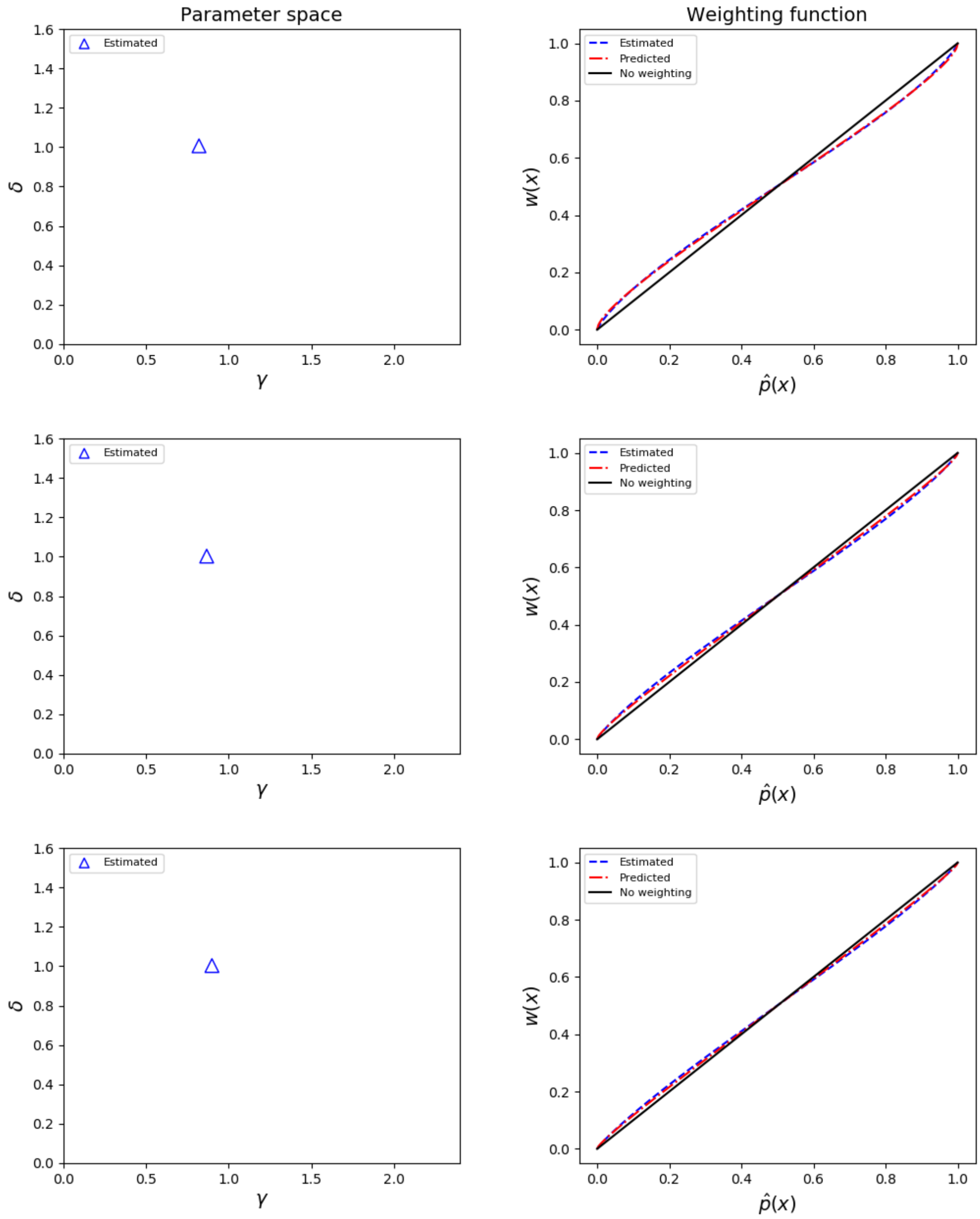
Probability Weighting function for LML-Agent 7



Probability Weighting function for LML-Agent 8



Probability Weighting function for LML-Agent 9



Probability Weighting function for LML-Agent 10