American Psychological Association

Journal of Experimental Psychology: General

0096-3445/18/$12.00

2018, Vol. 147, No. 4, 462–484
http://dx.doi.org/10.1037/xge0000403

Taxing Cognitive Capacities Reduces Choice Consistency Rather Than Preference: A Model-Based Test

Sebastian Olschewski and Jörg Rieskamp
University of Basel

Benjamin Scheibehenne
University of Geneva

How do people make preferential choices in situations where their cognitive capacities are limited? Many studies link the manipulation of cognitive resources to qualitative changes in preferences. However, there is a widely overlooked alternative hypothesis, namely, that a reduction in cognitive capacities leads to an increase in choice inconsistency. We developed a mathematical model and followed a hierarchical Bayesian estimation approach to test to what extent a reduction in cognitive capacities leads to a shift in preference or an increase in choice inconsistency. Using a within-subject *n*-back task to manipulate cognitive load, we conducted three experiments across different choice domains: risky choice, temporal discounting, and strategic interaction. Across all three domains, results show that a reduction in cognitive capacities predominantly affected participants' level of choice consistency rather than their respective preference. These results hold on an individual and a group level. In sum, our approach and the mathematical model we used provide a rigorous and general test of how reduced cognitive capacities affect people's decision-making.

Keywords: choice consistency, preferential choice, stochastic choice, cognitive load, economic choice

In many situations people are distracted, stressed, tired, or occupied with several things at the same time. These characteristics are often not matched in laboratory studies in which participants are able to direct all their attention to the task given by the experimenter. Therefore, researchers in psychology and economics have recently tried to better understand if and how a reduction in cognitive capacities affects behavior across a wide range of tasks. Prominent examples include research on risky (economic) choices (Benjamin, Brown, & Shapiro, 2013; Deck & Jahedi, 2015; Freeman & Muraven, 2010), trade-offs between short-term and long-term rewards (Deck & Jahedi, 2015; Ebert, 2001; Hinson, Jameson, & Whitney, 2003; Joireman, Balliet, Sprott, Spangenberg, & Schultz, 2008), and strategic interaction games (Cappelletti, Güth, & Ploner, 2011; Halali, Bereby-Meyer, & Meiran, 2014; Schulz, Fischbacher, Thöni, & Utikal, 2014). The present work examines how a reduction in cognitive capacities affects decision making across these three major preferential choice domains. Although the three domains are grounded in different theories and build on different hypotheses about which cognitive resources are required to solve a given task, there is a striking similarity in the basic assumption that reduced cognitive resources can lead to systematic changes in people's preferences regarding risk, time, or fairness.

There is, however, a plausible alternative hypothesis to a systematic preference shift: A reduction in cognitive capacities might lead to an increase in choice inconsistencies; for example, because people pay less attention to the stimuli, they are less precise in integrating the stimulus information, or they make more random choices when implementing their decisions. An increase in inconsistency can easily be mistaken for a systematic preference shift. For example, when diminished cognitive resources lead to a higher probability of choosing an (unhealthy) cheesecake over a (more healthy) fruit salad (Shiv & Fedorikhin, 1999), it might be due to a genuine change in preference for the immediate (unhealthy) reward but it might also be due to an increase in inconsistency: For a person who usually chooses the healthy fruit, higher rates of inconsistency will inevitably lead to a higher probability of choosing the inferior alternative (i.e., the unhealthy food).

An increase in choice inconsistency is a plausible alternative hypothesis when taking into account two closely related lines of research. The first is the effect of cognitive load in the domain of reasoning and problem solving in general (De Neys, 2006; Law, Logie, & Pearson, 2006; Meiser, Klauer, & Naumer, 2001; Phillips, Gilhooly, Logie, Della Sala, & Wynn, 2003). In these studies, participants had to solve math or logic problems while cognitive capacities were taxed with a secondary task. Performing a secondary task increased the number of errors committed or reduced the number of problems solved compared to a baseline condition. Just

Sebastian Olschewski and Jörg Rieskamp, Center for Economic Psychology, Department of Psychology, University of Basel; Benjamin Scheibehenne, School of Economics and Management (GSEM), University of Geneva.

This research was supported by a grant from the SNF: 100014_149846 to Jörg Rieskamp and Benjamin Scheibehenne. Raw data, mathematical model description, and data analysis code are available at the open science framework: osf.io/vfmt8. Parts of the data were presented at the SPUDM 2015, SJDM 2015, and the Thurgau Experimental Economic Meeting 2016. We thank Sarah Kuhn for helping with the data collection as well as Laura Wiles and Anita Todd for editing the article.

Correspondence concerning this article should be addressed to Sebastian Olschewski, Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland. E-mail: sebastian.olschewski@unibas.ch

as a reduction in cognitive capacities can impair participants' abilities to solve problems, it could also impair participants' performance in preferential choices. The second line of research has explored the link between intelligence or general cognitive abilities and preferential choices in correlative studies. Here a similar debate exists on whether general cognitive abilities are linked to preferences such as risk aversion or temporal discounting (Burks, Carpenter, Goette, & Rustichini, 2009; Dohmen, Falk, Huffman, & Sunde, 2010; Shamosh et al., 2008). Andersson, Tyran, Wengström, and Holm (2013) used a sophisticated experimental design to show that higher cognitive abilities can lead to both more or less risk taking. Thus, the authors concluded that cognitive abilities are related to choice consistency rather than to systematic differences in risk preference. These and other correlative findings make use of interindividual differences and usually use a mixture of cognitive skill measures. Although this does not provide a causal link between cognitive capacities and choice consistency, it further motivates the examination of cognitive load as a state manipulation of choice consistency.

Assessing choice inconsistencies in a preferential choice task requires assumptions about the choice process. Using a deterministic utility function to model preferential choices leaves no room for inconsistencies: People should always choose the option that maximizes (expected) utility. However, for a long time, researchers in decision making have emphasized that choices are not deterministic and that decision makers violate deterministic utility models on a regular basis (Mosteller & Nogee, 1951). One common solution in the risk literature is the application of a stochastic link function (Birnbaum & Bahra, 2012; Hey, 1995; Rieskamp, 2008; Wilcox, 2015).

A stochastic link function builds a bridge between deterministic utility models and the stochastic empirical nature of preferential choices. A *trembling hand* error, as an example of the fixed utility class, adds a certain probability of committing a choice error that means not choosing the option with the highest subjective utility (Harless & Camerer, 1994). Alternatively, random utility models, such as the *probit* choice models (Hey & Orme, 1994; Thurstone, 1927), assume that the utility of a choice option is not fixed but varies following a specific distribution. When making a choice, people will always choose the option with the momentarily larger utility; however, due to the variability of the utility, choice inconsistencies across many choices can result. The predicted choice probability of a random utility model is a function of the average utility difference of the considered choice options. Both fixed utility and random utility models predict that choices will vary across nearly identical choice situations (see Rieskamp, 2008). Therefore, for simplicity throughout this article, we will use the term *choice inconsistency* without preferring either of the two utility frameworks. In a preferential choice task, choice inconsistencies depend on the assumption of a given utility specification. Thus, in the context at hand, the consistency hypothesis states that reduced cognitive capacities increase the chance of choosing an inferior option, that is, an option with lower (average) utility as defined by the utility function.

We claim that refraining from a stochastic choice model—as is often done in studies of reduced cognitive capacities—can lead to unjustified conclusions. To avoid ambiguity and to determine whether the effect of a cognitive capacity reduction can be attributed to a shift in preference, an increase in inconsistencies, or both, we propose a general mathematical model framework. This framework maps onto different domains as well as different utility specifications. In each of the three domains we investigate, namely, risky choice, temporal discounting, and the ultimatum game, we use different utility functions to capture preferences in the respective domains and stochastic choice models to capture choice consistency. In this way, we demonstrate that our conclusions are generalizable across different domains, utility functions, and utility frameworks. We continue with a closer examination of decision-making research and cognitive capacity reduction manipulations in the respective areas.

Risky Choice

Risk-taking behavior has been assessed in a wide range of everyday behavior as well as experimental tasks (e.g. Charness, Gneezy, & Imas, 2013; Dohmen et al., 2011). A well-established way to measure risk preferences is to present choices between risky gambles that differ with respect to outcomes and outcome probabilities. For example, a choice could be between a sure option of receiving $10 or a risky option of receiving $15 with a probability of 75% and nothing otherwise. People commonly like high expected values of outcomes (i.e., returns) but do not like high variance of outcomes (i.e., risk; e.g., Pratt, 1964). By providing multiple pairs of gambles with different expected values and variances, the decision maker has to make repeated trade-offs between expected values and variances, which allows an estimation of individual utility functions, thereby characterizing people's risk attitudes. In general, the more concave the utility curvature, the more risk averse a person is.

One way to model a concave utility function is with a power function:

$$U(x) = x^{\beta}, \tag{1}$$

where x is the objective outcome and β the subjective risk preference parameter. β values below 1 lead to a concave utility function representing risk aversion and β values above 1 lead to a convex utility function and hence risk-seeking behavior. The power utility function has been rejected on empirical grounds many times, which led to the development of rank-dependent utility models, of which arguably the most prominent is cumulative prospect theory (Tversky & Kahneman, 1992). Cumulative prospect theory also makes use of a power utility function, but it adds an editing phase to distinguish gains from losses as well as the assumption of loss aversion and probability weighting. Using only gambles in the gain domain, cumulative prospect theory as originally stated in Tversky and Kahneman (1992) has just one more parameter than the power utility function. This parameter governs the weighting function that transforms probabilities into subjective decision weights as follows:

$$W(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{(1/\gamma)}}. \tag{2}$$

Finally, a different way to model risk preferences is to assume a linear utility function but to introduce a bias term (Stewart, Reimers, & Harris, 2014). The core of this idea can be traced back to mean-variance models in the financial literature that were shown to approximate concave utility functions under certain assumptions (Levy & Markowitz, 1979). Here we further simplified the model by assuming equal variance differences between gambles (for details see the Method section). This leads to an expected value model with one free parameter that captures a choice bias for the riskier or safer of two options,

$$EV_{\text{riskier}} - EV_{\text{safer}} + \beta. \tag{3}$$

Typical studies in the domain of reduced cognitive capacities in risky choice, however, rarely estimate utility functions (e.g., Benjamin et al., 2013; Deck & Jahedi, 2015; Freeman & Muraven, 2010). For example, in the study by Deck and Jahedi (2015), participants repeatedly chose between risky gambles and safe outcomes. The risky gamble was a 50–50 chance to win either a high or a low amount of money and the alternative safe outcome had an expected value in between these two outcomes. Similar decisions were also made in the loss domain. The authors manipulated cognitive capacity with a dual-task design, in which participants had to remember either a one-digit (low cognitive load) or an eight-digit (high cognitive load) number during each choice. The observed data indicated that on average across all participants in the gain domain, cognitive load significantly decreased the choice share of the risky gamble over the safe option from 59.5% to 52.7%. Here, as in other studies mentioned above, our observation about the ambiguity of the reported effect applies: In line with the interpretation of the studies' authors, the data can be explained as a genuine shift in risk preferences, but alternatively an increase in choice inconsistencies can account for observed choice proportions closer to a random choice level of 50%. To resolve this ambiguity, both preference and choice consistency have to be assessed conjointly, which can be done by applying the quantitative model we present below.

Temporal Discounting

In general, people prefer immediate over delayed gratification, as can be seen by measured (implicit) discount rates (Frederick, Loewenstein, & O'Donoghue, 2002). From an economic perspective, it makes sense to discount future outcomes (Fisher, 1930), and the discount rate can be partly reflected in a market's interest rate. However, it has been experimentally shown that people sometimes act as if they were discounting future outcomes more strongly than the market interest rate would suggest. Such preferences for immediate rewards are often explained by impulsive behavior or self-control problems (O'Donoghue & Rabin, 2000). To elicit people's time preferences, it is common to let people choose between different monetary amounts that are received at different time points in the future. Here, people have to trade off between sooner smaller amounts and later larger amounts of money. In this paradigm, the (implicit) discounting rate is inferred by setting up a discounting function that is consistent with most choices. Likewise, it is possible to ask people directly for the present value of a certain amount that is received at a specific time point in the future. From the stated present value, a discounting rate can be determined that characterizes a person's time preference.

When dealing with monetary amounts that occur at different time points, economic theory prescribes an exponential discounting function as the normative standard (Samuelson, 1937),

$$\frac{outcome}{\exp(\kappa \cdot delay)}, \tag{4}$$

where the discounting factor κ represents the discounting of future outcomes, with larger values for κ implying stronger discounting and giving more weight to immediate outcomes. In contrast, other functions have been suggested to describe people's observed time preferences, with the one-parameter hyperbolic discounting (Ainslie, 1975) function as a prominent example:

$$\frac{outcome}{1 + \kappa \cdot delay}, \tag{5}$$

where κ has a similar interpretation as before. Psychologically, a larger κ in hyperbolic discounting is often interpreted as more impulsive behavior. In general, hyperbolic discounting often describes empirical time preferences better than exponential discounting (Frederick et al., 2002). More recently, several different and more complex discounting functions have been discussed on empirical, theoretical, or neuroscientific grounds (Ebert & Prelec, 2007; McClure, Ericson, Laibson, Loewenstein, & Cohen, 2007; Peters, Miedl, & Büchel, 2012). Here, as one representative of this class of models, we examined an alternative two-parameter hyperbolic discounting function from Green and Myerson (2004):

$$\frac{outcome}{(1 + \kappa \cdot delay)^{\sigma}}, \tag{6}$$

where κ again captures discounting and σ is a parameter that captures nonlinear scaling of the denominator. If the scaling parameter is smaller than 1, this implies weaker discounting compared with the one-parameter hyperbolic model.

Typically, studies examining time preferences under reduced cognitive capacities used repeated binary choices between immediate and delayed rewards (Deck & Jahedi, 2015; Ebert, 2001; Hinson et al., 2003; Joireman et al., 2008). In Hinson, Jameson, and Whitney (2003), participants made choices while under high cognitive load (i.e., remembering letters) or low cognitive load (i.e., pressing letters after each decision). The authors found that participants' hyperbolic discounting factors were larger under high compared with low cognitive load. Hence, the authors concluded that cognitive load leads to a shift in time preferences. However, again, it could also be that cognitive load increases choice inconsistency. This idea has been tested by Franco-Watkins, Pashler, and Rickard (2006), who reanalyzed the data of Hinson et al. (2003) and argued that there was no shift in time preference, but only an increase in inconsistencies that drag choice proportion closer to 50% (from 25% to 30%). This finding was further corroborated by an additional experiment of the same authors (Franco-Watkins, Rickard, & Pashler, 2010). To resolve this debate, it is necessary to rigorously test the preference-shift hypothesis against the choice-consistency hypothesis. This is possible with the mathematical model presented below.

Fairness Preference

Preference for fairness develops early in life, exists in human beings as well as in animals, and is claimed to have an important impact on the development of cooperation (Brosnan & de Waal, 2014; Knafo, Zahn-Waxler, Van Hulle, Robinson, & Rhee, 2008). Fairness preferences in economics and psychology are often studied in the domain of strategic interaction games. Typical examples are the ultimatum game (Güth, Schmittberger, & Schwarze, 1982) and the dictator game (Kahneman, Knetsch, & Thaler, 1986). In the ultimatum game, one person, the proposer, decides how to distribute a given outcome and the other person, the responder, can decide to accept or to reject the distribution. If the responder accepts, then both players get an outcome according to the proposed distribution, but if the responder rejects, both players get nothing. In the dictator game, again, one participant decides how to distribute a given amount of money between him- or herself and another person. However, the other person has no choice and only passively receives the distributed amount. Typically in these games, dictators choose to give a nontrivial share to the receiver, and most responders in the ultimatum game reject distributions that give less than 20% of the original outcome (Camerer & Thaler, 1995).

This unselfish behavior has often been explained by fairness preferences, according to which people do not care only about their personal monetary outcomes but are also concerned about the monetary outcomes for others. One way to model social preferences is to define a utility function that captures the personal monetary outcome but also the outcome for another person. This idea has been formalized by Fehr and Schmidt (1999) in their inequity aversion utility function, defined as

$$U(x, y) = x - \alpha \cdot \max(0, y - x) - \beta \cdot \max(0, x - y), \qquad (7)$$

where the utility U for a person is the sum of that person's own outcome, x, and the difference between that outcome and the outcome of another person, y. There are two free parameters: α, a measure of aversion to inequity disadvantageous to oneself or first-order inequity aversion; and β, a measure of aversion to inequity that favors oneself, or second-order inequity aversion. The authors claim that both types of inequity matter, but that second-order inequity aversion has less weight than first-order inequity aversion.

A slightly different specification of the same idea has been proposed by Bolton and Ockenfels (2000). Instead of the difference between a person's own and another person's outcome, they used the ratio. This specification allows for diminishing or increasing marginal disutility from unfair distributions:

$$U(x, y) = x - \alpha \cdot \max\left(0, \left(\frac{x}{x + y} - \frac{1}{2}\right)^2\right). \qquad (8)$$

Many studies examining fairness preferences under reduced cognitive capacities used either dictator or ultimatum games (Cappelletti et al., 2011; Halali et al., 2014; Schulz et al., 2014). Schulz, Fischbacher, Thöni, and Utikal (2014), for example, used 20 minidictator games, where participants had to choose between two different distributions. To manipulate cognitive load, the authors used a 0- or a 2-back task: Participants heard a sequence of letters and in the 0-back task had to press a button whenever a target letter was heard (control condition) whereas in the 2-back task they had to press a button when the currently heard letter was the same as the letter presented two places back (load condition). The authors found that the choice of the fair allocation increased from 30.9% to 43.3% from the control to the load condition. In line with other studies cited above, the authors concluded that under high cognitive load, participants' preferences shifted toward more fairness. Interestingly, Schulz et al. (2014) also reported that participants reacted more sensitively to the allocation alternatives in the control condition: In each trial, participants had to decide between an almost fair and an unfair allocation and the degree of unfairness was varied from a share of 60% up to 100% for the dictator. It was observed that dictators chose the fair allocation more often when the alternative allocation was very unfair, compared with cases where the unfair allocation was closer to the fair allocation. This effect was less pronounced for participants in the high load condition. In line with our reasoning, this finding could also be interpreted as a decrease in sensitivity or an increase in inconsistency under cognitive load. As in the previous domains, to rigorously test these two competing hypotheses, preference shifts and choice errors have to be assessed conjointly in a mathematical model.

Stochastic Choice Models

To account for the probabilistic character of preferential choices (see also Rieskamp, 2008) we add choice rules that lead to probabilistic choice predictions to the respective utility functions. We used two different choice models, namely, probit and trembling hand, to generalize over specific mathematical implementations and also different stochastic utility frameworks. According to random utility models the utility of an option varies. The probability of choosing an option is determined by the probability that one option has a higher utility than the other option. The probit random utility model assumes normally distributed utilities and can be decomposed into a stable and a random component:

$$U_{\text{stochastic}} = U(x) + \epsilon, \qquad (9)$$

with ϵ being normally distributed with mean 0 and variance θ and where $U(x)$ is the constant utility of the option. In case of choices, the probit transformation converts the preference order of different options into a probability of choosing the respective option. In case of valuations, answers are modeled as stemming from the respective normal distribution. The θ specifies the variability of the normally distributed utility around the stable utility prediction from a deterministic utility function. In general, the larger θ, the higher the observed choice inconsistencies and the more often an option with a lower mean utility is chosen.

The trembling hand model (Harless & Camerer, 1994) assumes that people have a fixed utility for each choice option, but when choosing between the two options they will perform an error with a constant probability and choose the inferior option. Assuming option y has higher utility to the decider than option x this means

$$prob(y) = \text{step}(U(y) - U(x)) \cdot (1 - \theta) + \text{step}(U(x) - U(y)) \cdot \theta, \qquad (10)$$

where the step function takes a value of 1 if $U(y) - U(x)$ is positive and 0 otherwise. In our example the probability of choosing y would be determined by the first term of the equation, that is, 1 minus the trembling hand error θ, whereas the second term would become 0. In the case of a valuation, a trembling hand error is modeled by drawing the valuation from two different distributions: from a normal distribution with the mean determined by a given utility model, or from a uniform distribution across the whole answer scale space. The first distribution corresponds to a valuation according to the utility model, whereas the second distribution represents a random valuation. The probability with which valuations are explained by a draw from the uniform distribution equals the trembling hand error as defined previously. Finally, the higher the trembling hand error, the higher the choice inconsistency.

Overview of Experiments

In sum, in very different domains of decision making, such as risk taking, intertemporal choice, and social interactions, researchers claim that reduced cognitive capacities lead to systematic shifts in people's preferences. To test the hypothesis of a change of preferences against the alternative hypothesis of an increase in choice inconsistency, we made use of the following methodological advancements: First, we used a within-subject design. This allows observation of the impact of reduced cognitive capacities on an individual level and thus a richer analysis than the usual comparisons of group differences. Second, we used mathematical models that explicitly incorporate a utility model and a stochastic choice theory, thus allowing for rigorous testing of the competing hypotheses. Third, we used choice and valuation tasks to generalize our findings across different response modes. Finally, by exploring three widely studied domains of decision making with a comparable mathematical modeling approach, we have corroborated the generalizability of our findings and have tried to unify methodology in studying preferential choices under reduced cognitive capacities.

We performed a model-recovery study to demonstrate that our modeling framework is able to identify and to distinguish between a shift in preferences and a shift in choice consistency and we refer to Appendix A for the details. Here, suffice to say that manipulating a preference parameter as well as manipulating choice consistency can be recovered with high power given a realistic set of parameter values and choice data. In addition, the discriminability is also very high, meaning that only rarely is a preference shift mis-specified as a shift in choice consistency and vice versa, given our modeling framework.

The idea of reduced cognitive capacities has been implemented in psychological experiments, for example, by inducing sleep deprivation or stress (e.g., Anderson & Dickinson, 2010; Morgado, Sousa, & Cerqueira, 2015). However, these designs are difficult to standardize and the duration and effect size of these manipulations over repeated choice situations might be questioned. Other researchers adopted sequential designs where participants were given a strenuous task prior to the task of interest (e.g., Freeman & Muraven, 2010). However, recently there have been doubts about whether this manipulation triggers a sensible effect size (Carter & McCullough, 2014). Taking these limitations into account, we used a simultaneous task design where people have a secondary task to perform while making preferential choices (cognitive load manipulation). More concretely, we induced cognitive load by means of an *n*-back task—a strong, reliable, and well-established manipulation (Cohen et al., 1994; Gevins & Cutillo, 1993; Kane, Conway, Miura, & Colflesh, 2007; Pashler, 1994).

In the following section we describe our first experiment in the domain of risky choice, starting with our mathematical model, and explain how the model captures preference shifts and changes in choice consistency. The second and third experiments on temporal discounting and strategic interactions follow. At the end, we discuss the converging evidence across all three experiments.

Experiment 1: Risk Taking

Method

Experimental design and mathematical model. In Experiment 1 we explored the effect of cognitive load on risky choices. Participants repeatedly chose between 160 binary two-outcome gambles presented on a computer screen. Each participant made half of the choices under cognitive load. In the load condition, participants performed an audio version of the *n*-back task in parallel with the main gamble task. In the control condition, a simplified version of the *n*-back task was presented (see below for details). The order of the manipulation was counterbalanced between participants. Because of the within-subject design, we analyzed the differences between the two conditions following a hierarchical Bayesian framework that captures individual- as well as group-level effects. In our main model, the subjective utility of a gamble is captured with a power function with the exponent β as a free parameter,

$$U_{.} = \Sigma_i^2 p_i \cdot x_i^{\beta+\delta_\beta \cdot cond}. \tag{11}$$

The utility difference between two gambles feeds into a probit choice function with one free parameter θ that measures the variability of the utility:

$$p_{risky} = \Phi\left(\frac{U_{\text{risky}} - U_{\text{safe}}}{\theta + \delta_\theta \cdot cond}\right), \tag{12}$$

where the difference between the control and load conditions within each participant is captured by a δ parameter introduced for both β and θ, governed by a dummy variable *cond* coded as -1 for the load condition and $+1$ for the control condition. Thus, the parameter values for the control and load conditions are calculated as adding or subtracting the respective δ from each average parameter value. This results in a composite measure for risk preference in the control, β_{control}, and the load, β_{load}, condition as well as a composite measure of error variance in the control, θ_{control}, and the load, θ_{load}, condition. Implemented this way, a difference in either preference or consistency between the control and the load condition will be reflected by the respective δ parameters. In particular, cognitive load could lead to a credible shift in preference; δ_β, to a credible shift in error variance; δ_θ, to a credible shift in both; or to no difference at all. The choice-consistency hypothesis states that cognitive load increases the inconsistency θ but leaves the risk preference β unchanged. Similarly, we instantiated the other models comprising of different utility functions and the trembling-hand choice rule as described in the introduction.

In all three experiments of the present work, we follow a two-step approach for inference: First, we tested the general existence of an effect of the cognitive load manipulation by model comparisons. We estimate WAICs for the full model and compare it with models that assume no effect of cognitive load (δs fixed to zero) and with models assuming just one single effect of either of the parameters. WAICs are established Bayesian model comparison tools and especially suited to hierarchical Bayesian modeling because they punish model complexity more accurately than comparable measures that just rely on the number of parameters (Vehtari, Gelman, & Gabry, 2016a, 2016b). In a second step, after testing the effect of the cognitive load manipulation, we followed an estimated approach for the parameters of the models to quantify and characterize the effects for the potential shifts in preferences and choice inconsistencies. Therefore, we examine the 95% highest density interval (HDI) of the posterior distribution of the differences in parameters between experimental conditions, δ. Differences where the HDI includes zero are not credibly different from zero, whereas differences that exclude the HDI are regarded as being credibly different from zero and thus it is concluded that such a parameter credibly differed due to the cognitive load manipulation (Kruschke, 2014).

We constructed a hierarchical Bayesian model along the lines proposed by Kruschke (2014). On an individual level we estimated four parameters for each participant (β_i, θ_i, $\delta_{\beta,i}$, and $\delta_{\theta,i}$). These parameters were drawn from a normal distribution with mean and standard deviation equal to the respective group-level parameters. Before being entered in the model, the parameters were further transformed as follows: For the parameters capturing preferences, δ_β was added to or subtracted from β depending on the experimental condition, and the sum was transformed into a uniform distribution between 0 and 3 by means of an inverse probit transformation that was scaled by the factor 3. Similarly, for the error variance, δ_θ was added to or subtracted from θ and the sum was transformed into a uniform distribution from 0 to 5. These transformed priors were chosen to be distributed in a broad range of plausible parameter estimates from previous estimation approaches in the literature or derived from theoretical considerations of the respective models. For all parameters, we made sure that the posterior estimates were not very close to the endpoints of a given range.

These transformations were done to facilitate the creation of intuitive and noninformative prior distributions on the group-level parameters. Means of all parameters were drawn from a normal distribution with a prior mean of 0 and a variance of 0.5. The variances of the group parameters were drawn from uniform distributions between 0 and 0.5. Given the sum of two parameters that are inverse probit transformed, a variance of 0.5 for each parameter guarantees that the transformed parameter combinations are truly uniform across the specified range. The priors for the alternative model specifications were constructed along the same principles.

This and all following data analyses were conducted with the JAGS package (Hornik, Leisch, & Zeileis, 2003) in RStudio (R Core Team, 2016; RStudio Team, 2015). WAICs were calculated from the likelihoods with the loo package (Vehtari et al., 2016a). All presented posterior estimates had an effective number of samples of at least 10,000 and were numerically approximated with three chains that mixed and converged, as indicated by the Gelman–Rubin statistic $\hat{R} < 1.02$ for all reported group posteriors (Gelman, Carlin, Stern, & Rubin, 2014). The source code of the models and the Bayesian analyses can be found at the Open Science Framework: https://osf.io/vfmt8/.

Gamble stimuli. The gambles in the choice task were presented adaptively (80 in each condition) to increase the efficiency of the experimental design. As a basis for the adaptive design, 400 pairs of risky two-outcome gambles were randomly created according to the following rules: Expected values were in the range of 40 to 100; the standard deviation of each gamble (i.e., its riskiness) ranged from 1 to 50; the pairs were equally distributed across 10 bins that varied in the expected value differences between the riskier and safer gamble (in ascending order). That is, in the first bin the expected value of the safer gamble was much higher than the expected value of the riskier gamble whereas it was the other way around in the tenth bin. Within each bin, the range of differences in standard deviation between the two gambles was similar so that the difference in standard deviation and in expected value were independent. The adaptive choice task itself consisted of several steps. In the first step, participants made 20 initial choices, based on two randomly selected gamble pairs from each bin. This was to guarantee that each participant made choices along the whole range of expected value differences. In a second step, an adaptive algorithm was implemented: First, each participant saw a pair of gambles from the fourth bin. Thereafter, when the riskier of the two gambles was chosen, the next pair of gambles was selected from a lower bin and vice versa whenever the safer of the two options was chosen.

The gambles in the second half of the experiment (i.e., either the load or the control condition) consisted of the same 80 gambles and were presented in the same order to make sure that the stimuli in the two conditions were comparable. Gambles were randomized with respect to their occurrence on the left or right side of the screen. Choices were self-paced and were made with the keys "D" for the left option and "L" for the right option. Figure 1 displays a screenshot of the experimental task. The software for this and the other experiments were programmed in PsychoPy (Peirce, 2007).

***N*-back task.** In parallel to the gamble task, participants also heard a continuous sequence of letters at 3-s intervals over earphones. Participants had to press the space bar on the keyboard whenever a target occurred. The definition of a target depended on the condition: In the load condition, a target occurred whenever the currently heard letter corresponded to the third latest letter in the sequence (hence, a 3-back task). In the control condition, every letter "L" represented a target. The control task did not require memory and thus should have put a significantly lower tax on cognitive capacities (Cohen et al., 1994; Miller, Price, Okun, Montijo, & Bowers, 2009). Participants had to press the button within 2,700 ms after the onset of the stimulus. In total there were eight different letters (D, F, H, L, K, N, P, R) and the sequences were randomly created with the constraint that 25% of a bundle of 40 consecutive letters contained a target. Feedback (right or wrong) was provided when the bar was pressed or when a target was missed. For every correct press of the space bar as well as for no reaction to nontargets, participants earned one point. To calculate the final score, the number of points was divided by the total number of letters heard.

Participants, incentives, and procedure. Forty psychology students (M_{age} = 24.23 years, SD = 5.47, seven male, 33 female) participated for course credit and a choice-dependent monetary bonus. Because we aimed for the analysis of our hierarchical Bayesian model framework, a traditional power analysis did not apply. Therefore, we opted for a convenient sample size of 40. The whole experiment lasted 60–90 min and participants earned on average 6.46 Swiss Francs (CHF; about $6.50) with a range of 2.00 to 13.80 CHF between participants. The experiment was approved by the institutional review board (IRB) of the psychology department of the University of Basel. Participants were welcomed at the laboratory, received written instructions, and gave informed consent. After the instructions, there were test questions to check whether they understood the decision task and the *n*-back task. It

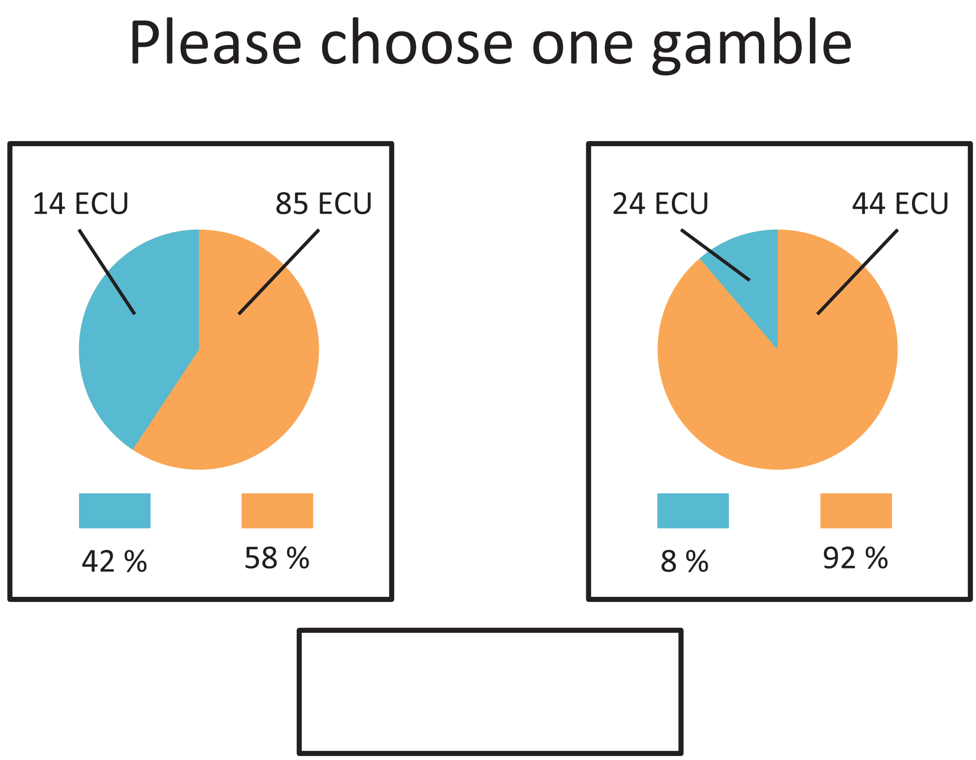

Figure 1. Schematic picture for one trial in the risky choice task. Participants chose one of the two gambles with the keyboard and heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = experimental currency unit; these were exchanged into Swiss Francs (10 ECU = 1 CHF). See the online article for the color version of this figure.

was guaranteed that only participants who answered all questions correctly started the experiment. If participants answered incorrectly, which happened only rarely, the instructor would reread the instructions and help the participant understand and produce the correct answer. The experiment was done in two blocks with one block containing the control task and one block the *n*-back task as secondary task. The order of the conditions was alternated between participants. As described, there were 80 self-paced gamble decisions in each block and there was a break of 10 min between the two blocks.

Both the decision task and the *n*-back task were incentivized. The decision task was incentivized by randomly selecting one of the gamble trials at the end of the experiment to play out. The outcome of the selected trial was then multiplied by the score from the *n*-back task. During the experiment all outcomes were shown in experimental currency units (ECUs), which were exchanged into Swiss Francs (10 ECU = 1 CHF). At the end of the choice task, participants performed a nonincentivized, automated version of the operation span (Ospan) task (Unsworth, Heitz, Schrock, & Engle, 2005). In this task, participants sat at the computer and solved math problems while having to remember up to seven letters, which they had to type in after a series of math problems. After this task participants were debriefed and paid.

Results

Descriptive results. Overall, participants chose the risky option 51.1% of the time in the control and 52.2% of the time in the load condition, Wilcoxon's test: $W(n = 40) = 361.5$, $p = .886$. Hence, the adaptive design managed to bring individual participants close to their respective indifference points. Yet, these percentages are hard to combine across participants, because everyone saw different gambles. Therefore, in Figure 2 we plot the percentage of risky choices across all participants separately for the control and load conditions and for different quantiles of expected value differences (calculated as the expected value of the riskier option minus the expected value of the safer option). The figure shows that the percentage of risky choices increased from the first to the fifth quantile, indicating that participants' choices were affected by the gambles' expected value. The figure further shows a visible difference between the control and load conditions: In the control condition, the increase in the percentage of risky choices is steeper than in the load condition. This is a first indication that choice consistency in the load condition was diminished. Reaction times were on average 5.8 s in the control and 6.6 s in the load condition. A *t* test across the individual log reaction time (RT) means showed no significant difference, $t(39) = -0.95$, $p = .347$.

Model results. Here we present the results for the full model as introduced in the Method section. The full model has a WAIC of 8,160. This is lower than the WAIC of a model fixing both δs to zero (8,253). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 8,191 or error with 8,179, respectively). This demonstrates that the full model, which assumes a shift in preferences and a shift in choice consistencies as

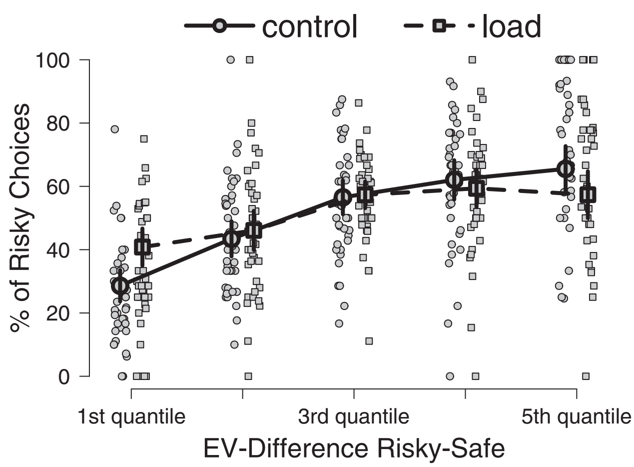

Figure 2. Experiment 1 risky gambles: Descriptive statistic for choice proportions for different quantiles of expected value (EV) differences between the riskier and the safer gamble with higher quantiles meaning higher EVs for the riskier gamble. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means, and error bars are 95% confidence intervals.

a result of the cognitive load manipulation is best to describe the data.

Now, the effect of cognitive load with respect to the different parameters is assessed and the estimates of risk preference β for the two conditions are presented in Figure 3. The estimated mean for the group-level posterior of the utility curvature parameter was $\beta_{control} = -1.23$ (*SD* = 0.15, 95% HDI [−1.53, −0.94]) in the control condition and $\beta_{load} = -1.13$ (*SD* = 0.21, 95% HDI [−1.54, −0.71]) in the load condition. Retransforming these values to the original scale, this corresponds to an average utility curvature parameter across both conditions of around 0.36. This means that the participants were overall quite risk averse. The group posterior for the difference in risk preference between the control and load conditions showed a mean of $\delta_\beta = -0.05$ (*SD* = 0.12). Because the posterior distribution overlaps 0 (95% HDI [−0.28, 0.19]), there is no credible difference between people's risk preferences in the two experimental conditions. Concerning the individual parameter estimates, it can be seen in Figure 3 that most participants scatter closely around the 45-degree line. There is also no trend of a majority of participants' risk preference parameter estimates increasing or decreasing, as can be seen by a binomial test (19 of 40 with $\delta_\beta > 0$, binomial test: $p = .875$). This means similar individual risk preferences between the control and load conditions, thus corroborating the group-level conclusion.

Figure 4 shows the posterior distribution of the utility variance on the group level θ, with a mean of $\theta_{control} = -2.04$ (*SD* = 0.07, 95% HDI [−2.18, −1.91]) in the control condition and a mean of an $\theta_{load} = -1.70$ (*SD* = 0.10, 95% HDI [−1.87, −1.50]) in the load condition. To put the absolute numbers into perspective, for an increase in terms of expected utility of 0.1 (outcomes were standardized to values between 0 and 1) at the switching point of the probit function, the percentage of choices for the riskier option increased from 50% to 83% on average in the control condition, but to only 67% in the load condition according to our choice model. This illustrates that participants under load were less sensitive to changes in expected value than in the control condition. The difference in the error variance parameter between the control and load conditions on the group level was $\delta_\theta = -0.17$ (*SD* = 0.05). This difference is credibly negative (95% HDI [−0.26, −0.09]). Thus, the utility variance was lower in the control compared to the load condition. This result is also corroborated on an individual level because all individual parameter estimates are above the 45-degree line shown in Figure 4 (or 40 of 40 participants had a $\delta_\theta < 0$).

Behavioral measures and robustness. In the *n*-back task, participants scored on average 84.46% correct, with a range from 69.5% to 92.9%. Because on average 25% of the stimuli were signals, never pressing a button would result in a score of 75%. Five participants earned below that score. To measure working memory capacity, we administered the automated Ospan task and report the total number of recalled letters: On average, participants achieved a score of 59.6 (range 34–75). Although we expected the individual differences in working memory capacity to explain some of the variance in the model parameters, there were no significant correlations between participants' Ospan scores and their estimated model parameters. There was also no significant correlation between the *n*-back score and the model parameters. Appendix B shows all correlations.

Finally, as mentioned above, we administered two alternative utility models: A linear utility model and cumulative prospect theory (see Equations 2 and 3). As is shown in Table 1, the two alternative models yield similar results. In both cases, choice inconsistency increased in the load compared with the control condition, whereas preferences remained unaltered. This is true in

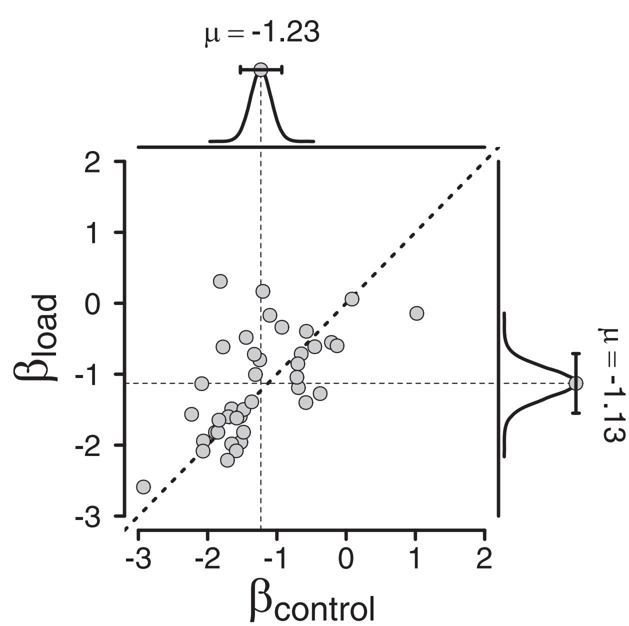

Figure 3. Experiment 1 risky gambles: Parameter estimates of risk preference β (on transformed scale): The *x*-axis shows individual risk preference parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of β in the respective conditions including the mean and the 95% highest posterior density interval.

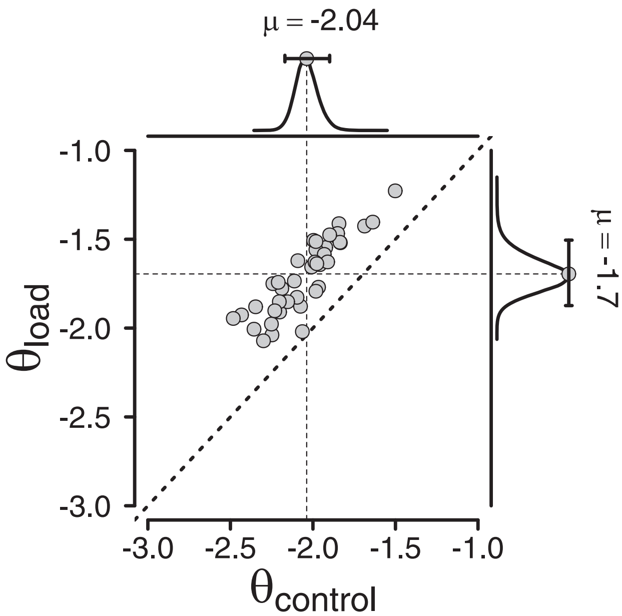

Figure 4. Experiment 1 risky gambles: Parameter estimates of choice sensitivity θ (on transformed scale): The *x*-axis shows individual choice sensitivity parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions including the mean and the 95% highest posterior density interval.

particular for cumulative prospect theory, showing that neither risk preference nor the probability weighting function is credibly influenced by the cognitive load manipulation. Using a different error model, namely, the trembling hand error (see Equation 10), leads to similar conclusions: Higher cognitive load increases the tremble error compared with the control condition for all tested utility models. From the WAIC scores, we conclude that the probit choice models describe our data on average better than the trembling hand error models. Although the prospect theory implementation did not show any differences in parameters between the two conditions other than the choice consistency parameter, it showed the best fits. To sum up, in line with the choice-consistency hypothesis, cognitive load led to an increase in choice consistency rather than a shift in risk preferences. This holds on an individual and on a group level. Furthermore, the results are consistent across three alternative utility models that are commonly used in the domain of risky choice and two different stochastic choice models.

Experiment 2: Temporal Discounting

Method

Experimental design and mathematical model. Experiment 2 tested how cognitive load affects temporal discounting of monetary outcomes. Participants were presented with different outcomes at different points in time (either one or two outcomes per trial) and had to state for how much money they were willing to sell their future outcome(s), otherwise known as their willingness to accept (WTA). Cognitive load was manipulated as a within-

Table 1
Experiment 1 Risky Gambles: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for All Model Specifications

	Linear utility			Power utility			Prospect theory			
Error model	Risk aversion	Sensitivity	WAIC	Risk aversion	Sensitivity	WAIC	Risk aversion	Weighting	Sensitivity	WAIC
Trembling hand	−.07 [−.22, .08]	−.20* [−.30, −.10]	8,338 [45]	−.14 [−.36, .06]	−.19* [−.29, −.09]	8,806 [77]	.01 [−.17, .19]	−.04 [−.18, .09]	−.26* [−.42, −.09]	7,962 [57]
Probit	−.08 [−.28, .12]	−.15* [−.24, −.07]	8,194 [49]	−.05 [−.28, .19]	−.17* [−.26, −.09]	8,160 [49]	.02 [−.17, .20]	−.04 [−.19, .09]	−.14* [−.24, −.04]	7,978 [56]

Note. For model specifications see introduction and the Method section of Experiment 1. Displayed are mean group posterior estimates and WAIC with 95% highest density intervals (HDI) or standard deviation of the WAIC in brackets.
* Significant differences between control and load condition according to the 95% HDI.

subject factor with an audio version of the 3-back task that was identical to the manipulation in Experiment 1.

To analyze the data we used a hierarchical Bayesian regression on the stated WTA prices. As our main model, we implemented the one-parameter hyperbolic discounting model with the discounting parameter κ specified as follows:

$$\mu_{\mathrm{dout}} = \frac{outcome}{(1 + \kappa + \delta_{\kappa} \cdot cond + \beta_{\mathrm{number}} \cdot number + \beta_{\mathrm{stake}} \cdot stake) \cdot delay}. \quad (13)$$

Here for a given trial, *outcome* stands for the monetary value in the experimental currency (see below) and *delay* is the amount of delay of the respective *outcome* in months. To present enough discounting trials to estimate the model parameters, we had to vary the stimulus characteristics with respect to the number of outcomes (either one or two delayed outcomes) and the stake (either low or high). To account for differences in discounting due to these factors, we included two additional dummies that were +1 for one-outcome trials or −1 for two-outcome trials, and +1 for high-stakes-outcome trials and −1 for low-stakes-outcome trials, respectively. The β parameters capture the corresponding effects.

To implement the probit choice model, the discounting function is fed into a Bayesian regression (Equation 14). This regression assumes a normal distribution around the discounted outcome, with a variance that equals the choice variability. The larger the variance, the broader the range of WTA prices for similar discounted amounts and the less sensitive the valuation with respect to the best fitting discounting parameter κ.

$$WTA \sim \mathrm{dnorm}(\mu_{\mathrm{dout}}, \theta + \delta_{\theta} \cdot cond). \quad (14)$$

Differences in parameter values between the two experimental conditions are captured by the δ terms as in Experiment 1. Again, the choice-consistency hypothesis states that cognitive load will change choice consistency but will leave time preference unaltered. The hierarchical Bayesian estimation was performed as in Experiment 1. The composite parameters, consisting of the main effect and the difference between the two experimental conditions, $\kappa_{\mathrm{control}}$, κ_{load}, $\theta_{\mathrm{control}}$, and θ_{load}, were set up uniformly from 0 to 0.2 and 0 to 1, respectively. We calculated $\theta_{\mathrm{control}}$ and θ_{load} on the precision scale, which transforms into the standard deviation scale as follows: $precision = 1/SD^2$. The exponential and the two-parameter hyperbolic discounting functions as well as the trembling hand choice rule were implemented accordingly, and results are presented at the end of the Experiment 2 Results section.

Temporal discounting stimuli. As stated above, there were two classes of stimuli to increase both the variety of the task and the number of informative trials: Some trials had only one delayed outcome and some trials had two outcomes that were paid out at different points in the future. All stimuli were created by defining 10 points in time ranging from 1 week to 1 year (0.25, 0.5, 0.75, 1, 1.5, 2, 3, 6, 9, 12 months). Then two ranges of outcomes were defined: low and high stakes. The low stakes ranged from 41 to 75 and the high stakes from 76 to 100 ECU either for one outcome or distributed over two outcomes. Finally, outcomes were randomly matched with the delay times. Forty stimuli each for one- and two-outcome trials were randomly selected and were identical for all participants. The task for the participants was to indicate their WTA, that is, their minimum selling price for each stimulus. They indicated their WTA with a slider that ranged from 0 to the undiscounted amount or the sum of undiscounted amounts (for the two-outcome trials) in that trial. Participants could move the slider until they were satisfied with its position and then confirmed their choice by clicking on the label with their current stated WTA price (see Figure 5).

Participants, incentives, and procedure. Forty-six psychology students (M_{age} = 22.2 years, SD = 5.1, seven male, 39 female) participated for course credit and a monetary bonus. The sample size was increase compared to the first study because participants had only half of the trials in this experiment. The whole experiment lasted 60 min to 75 min and participants earned on average 5.8 CHF (about $5.80; range 1.7–8.8 CHF). The experiment was approved by the IRB of the psychology department at the University of Basel.

Participants were welcomed at the laboratory, received written instructions, and gave informed consent. Only participants who correctly answered all questions concerning the experimental procedure could start the experiment (similar to the procedure for the first experiment). The experimental task was implemented on a computer in two blocks with a break of 10 min in between. In each of the two blocks, participants got 20 one-outcome and 20 two-outcome self-paced trials in randomized order. Whether participants started with the load or control block was alternated between participants.

Both the decision task and the *n*-back task were incentivized. At the end of the experiment, one of the trials was chosen at random and a Becker–DeGroot–Marschak auction was exercised (Becker, DeGroot, & Marschak, 1964): A random number between 0 and the maximum of the answer scale was drawn and compared to the participant's stated WTA for the given trial. If the random number was larger than or equal to the WTA, the participant's option for

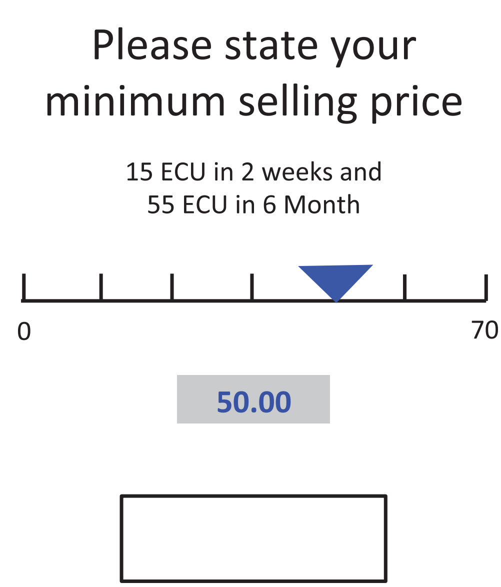

Figure 5. Schematic picture for one trial in the temporal discounting task. Participants dragged the blue triangle to their preferred value. The value chosen appeared in the gray rectangle below the scale. A choice was confirmed by clicking on the gray rectangle. Simultaneously, participants heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = experimental currency unit. See the online article for the color version of this figure.

gaining the future reward was sold for an immediate outcome proportional to the random draw. If the random draw was smaller, the participant's option was not sold and the participant kept the future outcome(s). The immediate or future outcome was then multiplied by the score of correct reactions to the *n*-back task (for calculation of the *n*-back score see Experiment 1). All shown outcomes were exchanged into Swiss Francs (5 ECU = 1 CHF). Immediate outcomes were paid in cash and delayed outcomes were wire-transferred at the respective times to the participant's private bank account. In the experiment, 12 participants received a random offer above their minimum selling price and thus received an immediate outcome. At the end of the two blocks, there was an unincentivized, computerized version of the Ospan task (Unsworth et al., 2005; see Method section of Experiment 1). After this task, participants were debriefed and paid.

Results

Descriptive results. Overall, participants selected an amount of 49.56 ECUs in the control and 49.22 ECUs in the load condition, which corresponds to an average discounting rate over all time intervals of 31.13% and 31.43%, respectively, Wilcoxon's test: $W(n = 46) = 554$, $p = .889$. A descriptive summary of the data shows the percentage of discounting indicated by the WTA prices across all participants separately for the control and load conditions and for different quantiles of delay (see Figure 6). The delay of the future outcome increased from the first to the fifth quantile and was in the case of two-outcome trials the average of the two delays. As expected, there was an overall trend of discounting increasing from the first to the fifth quantile. This shows that participants' WTA prices were affected by the (average) delay of the outcome(s). However, the increase in discounting for longer delayed outcomes was not very large, mainly due to the two-outcome trials where the average delay was less important in

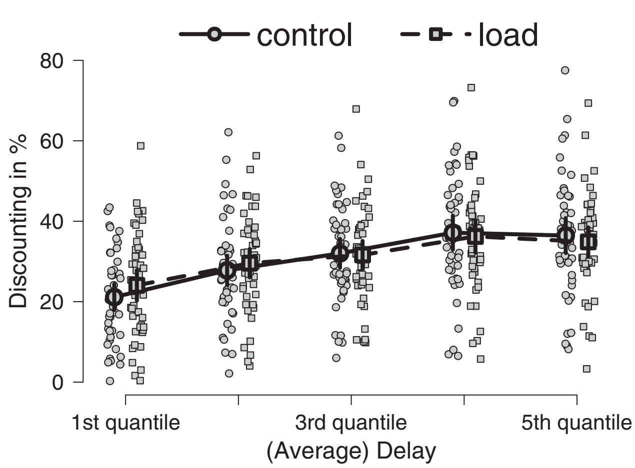

Figure 6. Experiment 2 temporal discounting: Stated willingness-to-accept prices are transferred into discounting percentages from the outcome or sum of outcomes in each trial. Discounting percentages are plotted for different quantiles of delay with higher quantiles meaning longer delays. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means and error bars are 95% confidence intervals.

determining WTA prices. Reaction times were on average 11.28 s in the control and 13.46 s in the load condition. A *t* test based on the individual log RT means did not show significant differences between the two conditions, $t(45) = -1.45$, $p = .154$.

Model results. Here we present the results for the full hierarchical Bayesian regression with hyperbolic discounting as introduced in Equation 14. The full model has a WAIC of 27,747. This is lower than the WAIC of a model fixing both δs to zero (27,969). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 27,855 or error with 27,874, respectively). This demonstrates that the full model, which assumes a shift in preferences and a shift in choice consistencies as a result of the cognitive load manipulation is best to describe the data.

To assess the magnitude of the effect on the preference and choice consistency parameters, Figure 7 first shows group and individual posterior estimates for the discounting parameter κ of the hyperbolic discounting model for the control and load conditions. Overall, the group posterior of the discounting parameter had a mean of $\kappa_{\text{control}} = -1.06$ ($SD = 0.08$, 95% HDI [−1.21, −0.91]) in the control condition and $\kappa_{\text{load}} = -1.02$ ($SD = 0.08$, 95% HDI [−1.17, −0.87]) in the load condition. Retransforming the average parameter estimates across both conditions results in a discounting rate of 0.15 for 1 month; hence, people showed considerable discounting behavior. As an example, $100 in 1 year is worth only $36.50 today, assuming hyperbolic discounting with the here-estimated discount rate. As the effect of our experimental manipulation on time preference, we estimated $\delta_\kappa = -0.02$ ($SD = 0.03$, 95% HDI [−0.07, 0.04]). Because 0 is included in the group-level distribution, we can conclude that there is no credible effect of the cognitive load manipulation on the discounting parameter. This is further corroborated by the estimates of individual parameters varying unsystematically between the control and load condition (21 of 46 participants with steeper discounting in the load condition $\delta_\kappa < 0$, binomial test: $p = .659$).

The group posterior of the parameter capturing an effect of two-outcome trials compared to one-outcome trials had a mean of $\beta_{\text{number}} = -0.39$ ($SD = 0.05$) and the 95% HPD interval excludes 0 95% HDI [−0.50, −0.29]. This means that two-outcome stimuli were more strongly discounted than one-outcome stimuli. This may be due to the discomfort of two different points of payment. Moreover, discounting was weaker in trials with high-stakes outcomes than in trials with low-stakes outcomes: The parameter capturing the effect of high- compared with low-stakes trials had a significant influence on discounting behavior ($\beta_{\text{stake}} = -0.04$, $SD = 0.02$, 95% HDI [−0.07, −0.01]).

The group-level posterior of consistency of WTA prices was measured with the precision θ (= 1/*variance*) of the normal distribution in Equation 14. Estimates are on a probit scale and results are shown in Figure 8. Neither the number of outcomes nor the amount at stake credibly influenced precision and thus no additional dummies were included. Estimation of the precision gives a group mean $\theta_{\text{control}} = -1.52$, $SD = 0.08$, 95% HDI [−1.68, −1.38] in the control condition and $\theta_{\text{load}} = -1.66$, $SD = 0.08$, 95% HDI [−1.81, −1.51] in the load condition. Retransformation of these values results in standard deviations of the WTA prices of 8.96 and 10.37, respectively. This shows that in the load condition, participants' WTA prices were more inconsistent with respect to the hyperbolic discounting model than in the control condition. In line with this, the results indicate a credible differ-

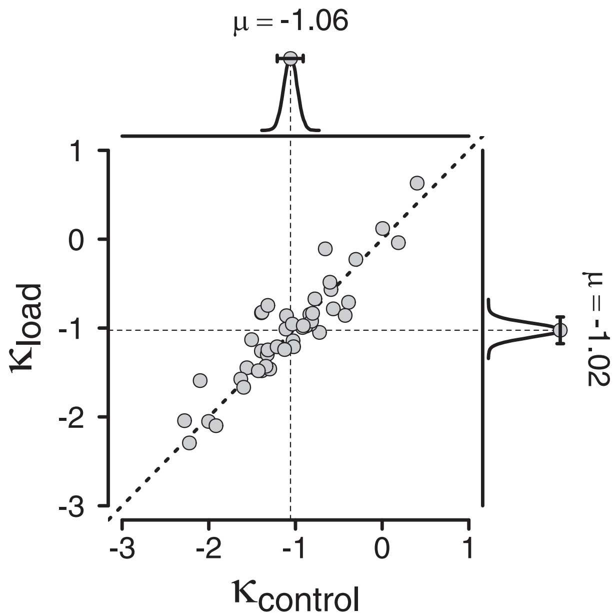

Figure 7. Experiment 2 temporal discounting: Parameter estimates of time preference κ (on transformed scale): The *x*-axis shows individual time preference parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of κ in the respective conditions including the mean and the 95% highest posterior density interval.

ence in the precision in the load compared with the control condition: The corresponding group-level posterior of the condition parameter had a mean of $\delta_\theta = 0.07$ ($SD = 0.02$) and the 95% HPD interval excludes 0 (95% HDI [0.02, 0.12]). This is corroborated on an individual level, as can be seen in Figure 8, where most participants' parameter estimates fall below the 45-degree line (35 of 46 participants $\delta_\theta > 0$, binomial test: $p = .001$).

Behavioral measures and robustness. In the *n*-back task, participants scored on average 84.62% correct, with a range from 74.40% to 94.29%. One participant's score was below the score that results if the button was never pressed. In the automated Ospan task, participants achieved an average score of 55.71, with a score range from 31 to 72. As in Experiment 1, there was no significant correlation between the Ospan measure or the *n*-back score and the model parameters (see Appendix B).

As a robustness check, we also implemented two alternative discounting models (see Equations 5 and 6): exponential discounting and a two-parameter hyperbolic discounting function proposed by Green and Myerson (2004). As shown in Table 2, for exponential discounting the choice inconsistency also increased in the load compared to the control condition. These results are robust to the use of a trembling hand error for the exponential as well as the one-parameter hyperbolic discounting model. For the two-parameter hyperbolic discounting function, two effects of the cognitive load manipulation were found: Both the discounting parameter and the choice consistency parameter differed credibly between the control and load conditions. Thus, this model specification cannot distinguish between an effect of preference or choice consistency. One reason might be that the two parameters affecting discounting of outcomes in this model (discounting and scaling) have been shown to be highly correlated (Peters et al., 2012). Consequently, although this model seems best in explaining the data taking the WAIC criterion, the additional mathematical complexity might make it more difficult to identify the source of the cognitive load effect.

In summary, these results indicate that cognitive load affected choice consistency rather than time preference, both for exponential and hyperbolic discounting and for two different choice rules. With a two-parameter hyperbolic discounting model, cognitive load seems to affect both preference and consistency. Overall, these findings accord with the results of the first experiment.

Experiment 3: Fairness Preferences

Method

Experimental design and mathematical model. In Experiment 3 we examined the influence of cognitive load on fairness preferences in social interactions. Participants in the experiment took the role of the responder in a sequence of one-shot miniultimatum games (Bolton & Zwick, 1995). As in the regular ultimatum game, the proposer distributes money between her- or himself and the responder. The responder can reject offers, in which case both participants get nothing. In miniultimatum games, the proposer can only decide between two given distributions, so we created different choice situations that allowed for repeated, nontrivial trials. Cognitive load was manipulated as a within-

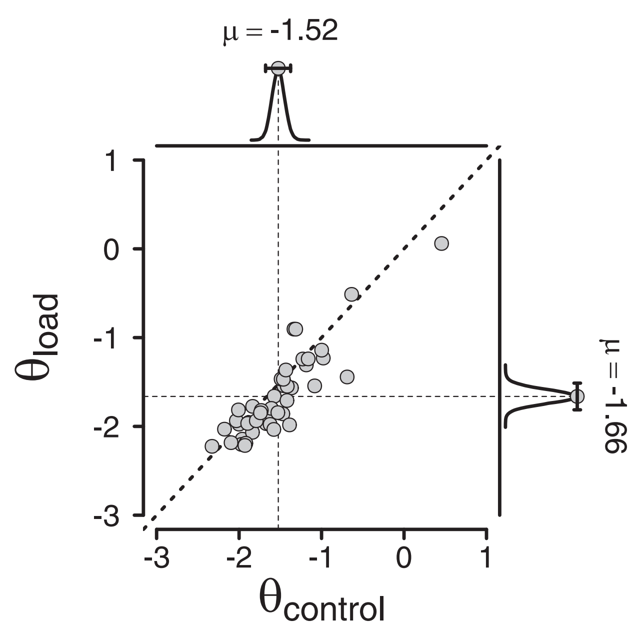

Figure 8. Experiment 2 temporal discounting: Parameter estimates of choice sensitivity θ (on transformed scale): The *x*-axis shows individual choice sensitivity parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions including the mean and the 95% highest posterior density interval.

Table 2
Experiment 2 Temporal Discounting: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for All Model Specifications

	Exponential			Hyperbolic 1			Hyperbolic 2			
Error model	Discounting	Sensitivity	WAIC	Discounting	Sensitivity	WAIC	Discounting	Scaling	Sensitivity	WAIC
Trembling hand	.00 [−.04, .03]	−.30* [−.52, −.07]	27,698 [98]	−.02 [−.07, .03]	−.34* [−.59, −.11]	27,261 [101]	−.17* [−.30, −.04]	.03 [−.02, .08]	−.30* [−.53, −.07]	26,525 [106]
Probit	−.01 [−.04, .03]	.05* [.01, .08]	28,436 [113]	−.02 [−.07, .04]	.07* [.02, .12]	27,747 [115]	−.16* [−.28, −.03]	.03 [−.02, .08]	.06* [.02, .10]	26,890 [120]

Note. For model specifications see introduction and Method section of Experiment 2. Displayed are mean group posterior estimates and WAIC with 95% highest density intervals (HDI) or standard deviation of the WAIC in brackets.
* Significant differences between control and load condition according to the 95% HDI.

subject factor with an audio version of the *n*-back task as in the previous experiments.

Our main model describes responders' rejection rates with a simplified version of the inequity aversion model from Fehr and Schmidt (1999) that takes only first-order inequity aversion into account. According to this model, we define the utility for a responder as

$$U_{\mathrm{dist}} = resp - (\alpha + \delta_{\alpha} \cdot cond) \cdot \max(0, prop - resp). \quad (15)$$

Here the rejection rates of responders depend on the amount the responder gets, *resp*, and the inequity against the responder is calculated as the difference between proposer and responder outcome or 0 if the responder gets more than the proposer. The parameter α is usually negative and estimates how important inequity is in determining the rejection rate and thus measures inequity aversion or fairness preference. We also estimated the inequity aversion utility function as specified in Bolton and Ockenfels (2000) and the full inequity aversion from Fehr and Schmidt (1999) including second-order inequity aversion (see the Robustness section below).

To account for choice consistency, our main model uses the probit formula with choice variability θ, similar to in the previous two experiments. In the context at hand, the probability of rejecting a given offer was specified as follows:

$$p(reject) = \Phi\left(\frac{0 - U_{\mathrm{dist}}}{\theta + \delta_{\theta} \cdot cond}\right). \quad (16)$$

As an example of how the model works, if the utility from the proposed distribution is negative (this happens given a certain α and if the proposer outcome is sufficiently larger than the responder outcome), the probability of rejecting the distribution is larger than 50%. This is similar when implementing a trembling hand error, where a free parameter captures just the probability of rejecting an offer with a positive utility and accepting an offer with a negative utility. Differences between the load and the control conditions are again captured by the δ terms for α and θ, respectively. As before, the choice-consistency hypothesis states that cognitive load changes the sensitivity parameter θ of the responder, but not the responder's inequity aversion α. The hierarchical Bayesian estimation is similar to in the previous experiments. The composite parameters capturing inequity aversion, $\alpha_{\mathrm{control}}$ and α_{load}, are set up uniformly between 0 and 5 and the parameters capturing error variance, $\theta_{\mathrm{control}}$ and θ_{load}, are set up uniformly in the range of 0 to 100. The additional model specifications are set up in the same way.

Miniultimatum game stimuli. In total we created 40 miniultimatum games. Responder outcomes ranged from 0 to 90 and proposer outcomes from 15 to 120 ECU. The sum of responder and proposer outcomes in each of the possible distributions was not necessarily equal. This means there were trade-offs between the overall outcome (i.e., social welfare) and the respective outcome distributions. All miniultimatum games were pretested to make sure that they provided nontrivial distribution options for the proposer and that they entailed nonnegligible rejection rates on the part of the responder (Fleischhut, Artinger, Olschweski, Volz, & Hertwig, 2014).

All participants saw the same 40 miniultimatum games, but in different conditions (control or load) and in a randomized order. Incentivized choices from five proposers were collected before the main experiment started. Proposer choices were only important for incentivizing responders and are therefore not reported here. Responder choices were collected with the strategy method. This means responders had to state their acceptance or rejection for each of the two alternative distributions in each miniultimatum game before they knew which distribution was chosen. This allowed us to elicit a larger amount of responder data. In each trial, participants saw the currently offered distribution in the middle of the screen and had to state whether to accept or reject with the letters "D" and "L." In the upper right corner of the screen they also saw the alternative distribution that the proposer could have chosen. Figure 9 shows a screenshot of the task.

Participants, incentives, and procedure. Fifty-seven psychology students (M_{age} = 25.07 years, *SD* = 8.06, 14 male, 43 female) participated for course credit and a choice-dependent monetary bonus. A higher sample size than in the first experiment was chosen because of the reduced number of trials for each participant. The whole experiment lasted between 45 min and 60 min and participants earned on average 7.60 CHF (about $7.60) aggregated over responder choices and the *n*-back task. The payments varied from 3.70 to 11.29 CHF across participants. The variation in payment mainly occurred because for 15 participants a trial was chosen for payment where they rejected an ultimatum offer, thus these participants earned nothing from the ultimatum game. The experiment was approved by the IRB of the psychology department at the University of Basel.

Participants were welcomed at the laboratory, received written instructions, and gave informed consent. Only those participants who answered all questions concerning the experimental procedure correctly started the computerized experiment. The experiment consisted of two blocks with a 10-min break between the blocks. Participants got 40 self-paced choices (20 miniultimatum games) where they could accept or reject distributions that were randomized in each block. The two different distributions of one miniultimatum game did not necessarily follow each other, but they always appeared in the same block. Whether participants started with the load or the control block was alternated across participants.

Both the decision task and the *n*-back task were incentivized. One of the decision trials was chosen at random and matched with one of the five proposers. For the payment, only the responder choice for the distribution actually chosen by the matched proposer mattered. If at this distribution the responder accepted, then she or he would earn a payout proportional to the responder outcome in that distribution. If the responder rejected this distribution, he or she would get nothing. The five proposers were matched equally often to responders and at the end of all experiments for each proposer one responder was randomly chosen to be payoff relevant. The proposers got their money after the experiment ended via personal collection or bank transfer. In contrast to the previous experiments, the performance in the *n*-back task was incentivized by multiplying the obtained score by 5 CHF. The resulting amount was then added to the payment from the miniultimatum game (for calculation of the *n*-back score see Experiment 2). This was done to guarantee that participants who rejected many distributions also had an incentive to achieve a good *n*-back score. All shown outcomes were transferred into Swiss Francs (10 ECU = 1 CHF). At the end of the two blocks, there was an unincentivized, computerized version of the Ospan task (Unsworth et al., 2005) as in the previous experiments. After this task, participants were debriefed and paid.

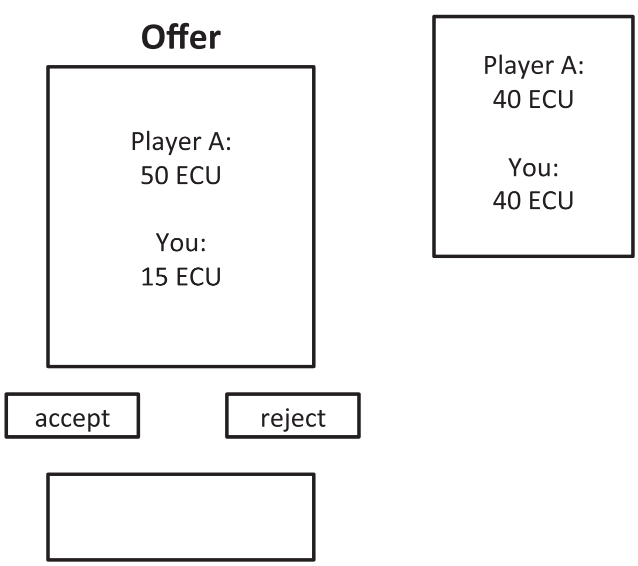

Figure 9. Schematic picture for one trial in the miniultimatum game. Participants chose whether to accept or reject an offer with the keyboard. A bit smaller on the upper right side of the screen, the distribution of the miniultimatum game that was not chosen was depicted. Simultaneously, participants heard letters over earphones; feedback on the auditive task was given in the blank rectangle below. ECU = experimental currency unit.

Results

Descriptive results. In the following we analyze only responder choices: On average participants rejected 37% of all offers, 36% in the control and 38% in the load condition, respectively, Wilcoxon's test: $W(n = 57) = 460$, $p = .130$. Figure 10 shows the average rejection rates for different levels of inequality separately for the control and load conditions. As expected, the rejection rate increased with the inequality of the distribution. At first glance, there is no visible difference between the control and load conditions. Participants took on average 2.82 s in the control condition and 4.81 s in the load condition to accept or reject an offer. Taking the means of all participants and calculating a paired *t* test, the difference in logarithmic RT between conditions is significant, $t(56) = -4.98$, $p < .001$.

Model results. Here, we present the results for the full model as introduced in Equation 16. The full model has a WAIC of 2,815. This is lower than the WAIC of a model fixing both δs to zero (2,989). In addition, it is also smaller than both models with one δ fixed to zero (either preference with 2,885 or error with 2,913, respectively). This demonstrates that the full model, which assumes a shift in preferences and a shift in choice consistencies as a result of the cognitive load manipulation is best to describe the data.

In the following the respective magnitudes of the effects of cognitive load on preference and choice consistency are assessed. Group-level posteriors of inequity aversion for both conditions as well as individual parameter estimates are depicted in Figure 11. Overall, inequity aversion α had a credible influence on choices

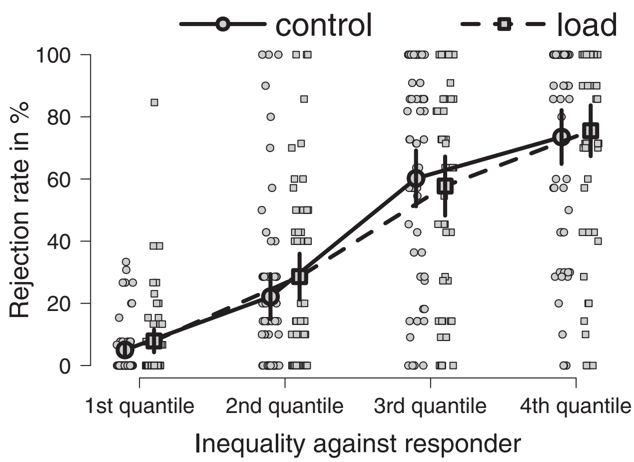

Figure 10. Experiment 3 miniultimatum game: Descriptive statistics of responder choices in the miniultimatum game. The rejection rates are plotted for different quantiles of inequality against the responder (outcome proposer minus outcome responder) with higher quantiles meaning higher inequity. Small dots and squares are individual choices in the control and load conditions, respectively. The larger dots and squares are group means, and error bars are 95% confidence intervals.

(α_{control} = −1.10, *SD* = 0.10, 95% HDI [−1.30, −0.89]; α_{load} = −1.03, *SD* = 0.11, 95% HDI [−1.23, −0.82]). Transformed on the original scale, overall inequity aversion equals 0.72. This reflects that higher inequity between proposer and responder outcomes leads to higher rejection rates (e.g., an offer of 25 for

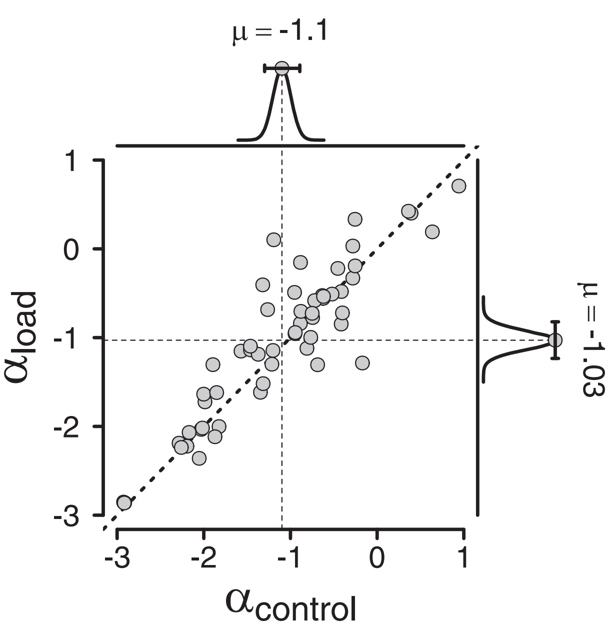

Figure 11. Experiment 3 ultimatum game: Parameter estimates of inequity aversion α (on transformed scale): The *x*-axis shows individual inequity aversion parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of α in the respective conditions, including the mean and the 95% highest posterior density interval.

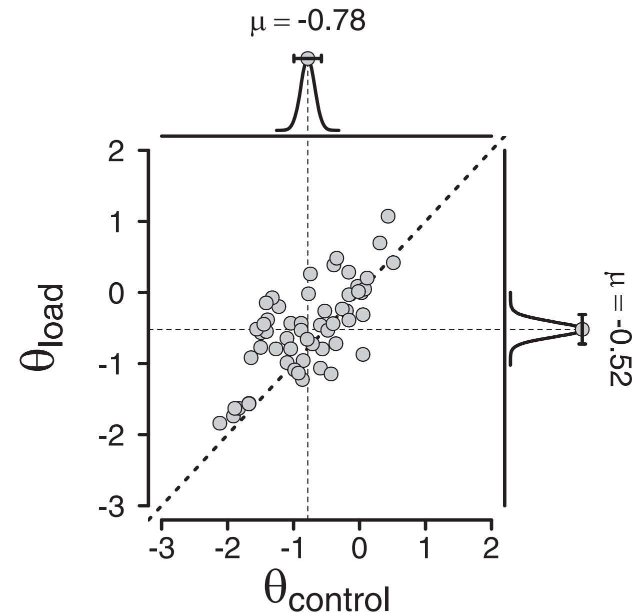

Figure 12. Experiment 3 ultimatum game: Parameter estimates of choice sensitivity θ (on transformed scale): The *x*-axis shows individual choice sensitivity parameter estimates in the control condition and the *y*-axis shows them in the load condition. Above and to the right of the plot are the group posterior distributions of θ in the respective conditions, including the mean and the 95% highest posterior density interval.

the responder while keeping 75 for oneself will be rejected most of the time according to the estimated inequity aversion). The difference between the control and load conditions is captured by δ_α. The group-level posterior distribution of δ_α is not credibly different from 0 (δ_α = −0.03, *SD* = 0.04, 95% HDI [−0.11, 0.05]. Thus, inequity aversion did not differ between the control and load conditions on a group level. In Figure 11, points below the 45-degree line signify lower values in the load compared with the control condition and vice versa for individuals. Here most points are very close to the 45-degree line, indicating that the cognitive load manipulation had no systematic effect on inequity aversion on the individual level. For 36 of 57 participants, inequality aversion was stronger in the load compared with the control condition (or $\delta_\alpha > 0$, binomial test: p = .063).

Looking at the choice variability θ, the group-level posterior means for both conditions are θ_{control} = −0.78 (*SD* = 0.11, 95% HDI [−0.99, −0.58]) and θ_{load} = −0.52 (*SD* = 0.11, 95% HDI [−0.73, −0.31]). To put these values into perspective, we retransformed the parameters to the variance scale, which results in 21.77 in the control and 30.15 in the load condition. Given the outcomes presented, this would mean that an increase in utility from 0 to 10 decreases the likelihood of rejection from 50% to 32% in the control condition and to only 37% in the load condition. The difference in the error variance between the control and load conditions was captured by δ_θ = −0.13 (*SD* = 0.06, 95% HDI [−0.24, −0.02]). Because the 95% HPD interval excludes 0, the choice inconsistency is credibly higher in the load compared to the control condition. Figure 12 shows the group-level posteriors for θ as well as the individual θs separately for the control and load conditions. Points above the 45-degree line mean higher inconsistencies in the load compared to the control condition for individuals. As can be seen, a majority of participants are above the 45-degree line, or for 39 of 57 participants, choice error was higher in the load compared to the control condition (i.e., $\delta_\theta < 0$, binomial test: $p = .008$).

Behavioral measures and robustness. In the *n*-back task, participants scored on average 83.85%, with a range from 68.35% to 95%. Five participants had scores below the guessing rate of 75%. In the automated Ospan task, participants achieved an average score of 55.98 (range 13–75). As in Experiments 1 and 2, there was no significant correlation between the Ospan measure or the *n*-back score and the model parameters (see Appendix B).

To check for robustness of our results, we administered two alternative other-regarding utility functions: the inequity aversion utility function proposed by Bolton and Ockenfels (2000) and the full Fehr and Schmidt (1999) utility model with an additional parameter for second-order inequity aversion (see Equations 7 and 8). Table 3 shows that for both alternative utility models, choice inconsistency increased in the load compared to the control condition, thus confirming the previous results. When using a trembling hand choice rule, the results also point in the same direction (i.e., less choice error in the control compared to the load condition), but the parameter differences are no longer significant. For the trembling hand error, there is also no effect of cognitive load on the inequity preference parameters in any of the three utility models. WAICs show consistently a worse fit of the trembling hand compared to the probit models. This indicates that the effect of cognitive load on responder behavior in the miniultimatum game is better captured by a choice model taking numerical utility differences into account than by a choice model that discards this information.

To conclude, we found that cognitive load affected choice consistency rather than fairness preference in the miniultimatum game on both an individual and a group level. This effect is robust to different other-regarding utility functions, but fails to show significant differences when using a trembling hand error model.

General Discussion

To test if a reduction in cognitive capacities leads to qualitative preference shifts, systematic increases in choice consistency, or both, we conducted three experiments across different domains of preferential decision making, including risky choice, temporal discounting, and strategic interaction. Across all three experiments, cognitive capacity was manipulated within subjects by means of a dual-task paradigm where participants completed an auditive 3-back task while making choices. A comparison of the hierarchical Bayesian models based on WAICs showed that the current choice data was described best when assuming changes in both, choice consistency and preference. A more thorough analysis based on estimating and comparing the models' parameters revealed that a reduction of cognitive capacities predominantly led to more choice inconsistencies rather than qualitative preference changes. These results hold on the group and individual level alike and are robust to the alternative model specifications that we tested. Furthermore, the results hold both for binary choices (Experiments 1 and 3) and economic valuations (Experiment 2). Thus, an increase in choice inconsistency as a result of a reduction in

Table 3
Experiment 3 Ultimatum Game: Mean Group Posterior Estimates of the Effect of Cognitive Load and WAICs for All Model Specifications

Error model	Fehr and Schmidt			Bolton and Ockenfels			Fehr and Schmidt Full			
	Inequity aversion I	Sensitivity	WAIC	Inequity aversion I	Sensitivity	WAIC	Inequity aversion I	Inequity aversion II	Sensitivity	WAIC
Trembling hand	−.04 [−.12, .03]	−.06 [−.15, .02]	3,030 [85]	−.62 [−1.82, .44]	−.06 [−.15, .03]	3,113 [84]	−.04 [−.11, .04]	−.01 [−1.39, 1.37]	−.06 [−.15, .02]	3,034 [85]
Probit	−.03 [−.11, .05]	−.13* [−.24, −.03]	2,815 [79]	−.13 [−1.11, .88]	−.14* [−.25, −.02]	2,945 [80]	−.03 [−.11, .05]	−.06 [−.56, .45]	−.11* [−.20, −.01]	2,758 [77]

Note. Inequity aversion I refers to first-order inequity aversion and Inequity aversion II refers to second-order inequity aversion as defined in the Introduction. For further model specifications see introduction and the Method section of Experiment 1. Displayed are mean group posterior estimates and WAIC with 95% highest density intervals (HDI) or standard deviation of the WAIC in brackets.
* Significant differences between control and load condition according to the 95% HDI.

cognitive capacities seems to be so far an underappreciated effect in preferential choice.

Our results provide a possible alternative explanation for recent studies claiming that a reduction in cognitive capacities can lead to preference changes. We argue that such a reduction, above all, increases choice inconsistency and that ignoring this effect can lead to a biased conclusion that could look like changes in preferences. Yet, unlike the studies cited in the introduction, the current studies did not find differences in choice proportions or valuation between the control and load conditions. This difference might be explained by the stimulus environment. The choice proportions, for example, in the risky choice experiment were very close to 50% for the safer and the riskier option in the control condition. In such an environment, an increase in choice error changes choice proportions symmetrically, whereas if choice proportions are lower or higher, an increase in choice error can drag choice proportions closer to 50% in a nonsymmetrical way. Moreover, a valuation task, as was used in the temporal discounting task, might not, in general, be prone to the problem of confusing choice consistency with preference shifts, because choice inconsistencies in the valuation task are arguably symmetrical (except for valuations at the margins). In sum, an observed group difference between control and load conditions might be mainly due to biased choice stimuli in the control group, and a modeling approach can help show the underlying effect of cognitive load on economic choices and valuations regardless of the stimulus environment.

Evidence for this view comes from the parameter recovery study (see Appendix A). Here, we manipulated choice proportions either as being at 50% (as in the risky choice experiment) or as deviating from 50%. In line with our reasoning above, differences in choice proportions due to a shift in choice consistency were only observed frequently when choice proportions in control deviated from 50%. Yet, even when a significant shift in choice proportions occur, our modeling framework can still distinguish between a shift in preferences and a shift in choice consistency.

To corroborate this simulation-based result, we also applied our model to a previously published data set that manipulated cognitive load and found a difference in choice proportion (Hinson et al., 2003). As described in the section about temporal discounting, Franco-Watkins et al. (2006) have previously reanalyzed the data and found evidence for an increase in choice inconsistencies. Using our modeling framework with hyperbolic discounting as the original authors, but—unlike them—adding a probit choice function to it and estimating the parameters in a hierarchical Bayesian approach, we conclude that the difference between the discounting parameters in the two conditions is not credibly different from 0 ($\delta_\kappa = -0.06$ $[-0.15, 0.02]$), whereas the difference in choice consistency differs from 0 ($\delta_\theta = -0.10$ $[-0.17, -0.02]$). Thus, unlike the authors of the original article but in line with the analysis of Franco-Watkins et al. (2006) our modeling approach shows that the observed effect is predominantly due to a shift in choice inconsistency and not a shift in preferences. This shows that the presented modeling approach can distinguish between shifts in choice inconsistencies and shifts in preferences even when the observed choice proportions differ between the control and the load condition. Thus, especially when choice proportions in the control condition deviate from 50%, merely analyzing differences in choice proportions or modeling the data without accounting for choice inconsistencies is not sufficient to detect genuine preference shifts.

Changing Preferences

Why should preferences change due to a reduction in cognitive capacities, as claimed by many recent studies described in the introduction? One possibility is that people become more impulsive or less self-controlled following a reduction in cognitive capacities. This idea is especially popular in food choice and temporal discounting studies (Hinson et al., 2003; Shiv & Fedorikhin, 1999). Here, it is argued that the affective or impulsive choice differs from the rational one (e.g., money now vs. tomorrow or cake vs. fruit). However, in other domains it is less obvious what an impulsive or rational choice could be: It appears plausible that more impulsive behavior in risky decision making corresponds to more choices of the riskier option. Likewise, impulsive behavior might also imply higher rates of rejection of unfair offers in strategic interactions. However, these conjectures appear less credible and opposite predictions are conceivable.

A more general explanation of why preferences might change is that cognitive load leads to qualitative changes in the underlying decision strategy. From the perspective of adaptive decision making (Payne, Bettman, & Johnson, 1993), people could apply an expected utility maximization strategy under full cognitive capacities, whereas with reduced cognitive capacities they might switch to simpler and cognitively less demanding choice strategies or heuristics. By ignoring some information and using less integrative steps, such a switch in strategies could systematically change choice behavior and thus parameters capturing preferences.

How general are our findings given the specific manipulation of cognitive load we used? The manipulation of cognitive capacities using an auditive 3-back task is thought to exert a high cognitive load (for a 2-back task as high cognitive load, see Perlstein, Dixit, Carter, Noll, & Cohen, 2003). Yet, weaker or even stronger manipulation of cognitive capacities could be used (e.g., a 2- or a 4-back task). Could a weaker or stronger manipulation lead to preference changes that we did not observe in our studies? A very mild manipulation would most likely fail to restrict at least some people's working memory, such that some people would not change their behavior at all. In contrast, when using an extremely strong manipulation it appears likely that people would have a hard time expressing any valid preference. In the most extreme case, choice consistency would be at a minimum and response behavior would be completely random. In between these extremes, however, there might be levels of cognitive load that lead to strategy switches. One could imagine that a steady increase of working memory load would increase choice inconsistency with a given strategy to the point where performance is so bad that people switch to a less demanding strategy that reduces inconsistencies compared with the more complex strategy (for the adaptivity of such a switch under time pressure, see the simulation in Payne, Bettman, & Johnson, 1988). On the other hand, Worthy, Otto, and Maddox (2012) showed that people changed their learning strategies in a dynamic decision-making task when under cognitive load, yet not in an adaptive way. This means strategies were different in a control compared to a load condition, regardless of the reward structure in the environment.

Such strategy-shift analyses need to identify plausible strategy shifts before the experiment in order to design stimuli that distinguish between strategies. Given the plethora of heuristics and strategies in the three preferential choice domains examined here, it is beyond the scope of our analysis to examine strategy shifts exhaustively. Yet, we deem it a very interesting question to examine in further studies whether there can be strategy shifts due to cognitive load, whether there are similarities in these shifts across different domains, and whether these shifts are adaptive. In the current study, we used a widely accepted manipulation to reduce cognitive capacities and a variety of standard utility-based choice models in different domains to conclude that changes in behavior could be better explained by an increase in inconsistencies than a shift in preferences. However, in principle this increase in inconsistency could also be explained by a qualitative strategy shift.

Cognitive Capacity Reductions More Broadly

Although we used a simultaneous task design, the current findings might also be relevant for situations where cognitive capacity is reduced through different means, for example, in a sequential task design: In a study by Freeman and Muraven (2010), participants watched a mute video and had to rate the actress' facial expression (see Experiment 2). At the same time they saw common English words on the screen, which they explicitly had to ignore in the load condition (referred to as the "depletion" condition by the authors). In a second step of the experiments, all participants had to pump a (digital) balloon to earn money, facing an increasing risk of the balloon bursting and themselves receiving nothing (BART task). Analyzing the choice pattern, the authors concluded that the depletion task led to an increase in risk-seeking behavior. A common explanation for this effect is to assume a reduction in self-control due to performing the preceding task (the strength model or ego depletion, e.g., Baumeister, Vohs, & Tice, 2007).

Our results with simultaneous task manipulations offer a new hypothesis to consider: Rather than a systematic loss of self-control and hence a qualitative shift in preferences toward more risky choices, the observed behavior could also be (partly) due to an increase in choice inconsistency. This alternative explanation might also be interesting in light of the recent debate about the effectiveness of sequential task manipulations: Although a meta-analysis by Hagger, Wood, Stiff, and Chatzisarantis (2010) found an overall significant effect size, in a reanalysis Carter and McCullough (2014) came to the conclusion that the effect size cannot be distinguished from 0. Looking more closely at the dependent task in the sphere of choices, we see that some studies measured preferences (e.g., Freeman & Muraven, 2010; Joireman et al., 2008) whereas others measured, for example, susceptibility to the attraction effect (Masicampo & Baumeister, 2008). Taking our results into account, we would expect an effect of the manipulation on a choice consistency parameter (which could in general affect susceptibility to context effects), rather than on a preference parameter. Distinguishing these fundamentally different dependent measures in a mathematical model might lead to a better understanding of the effect of the sequential task manipulation on choices.

Finally, a more general approach could examine how decision errors and preference shifts are related to different classes of manipulations, such as sequential or simultaneous approaches. When considering that the two manipulations follow different theoretical constructs (working memory capacity vs. self-control), a systematic examination seems worthwhile. In addition, also time pressure, stress, and sleep deprivation arguably reduce cognitive capacities. For this more general perspective, our results make an important contribution to the study of reduced cognitive capacities in decision making as they offer a parsimonious explanation of effects reported in different areas (see Johnson, 2008).

Testing Stochastic Choice Rules

Whereas many axiomatic choice theories have neglected the stochastic element of choice (e.g., Von Neumann & Morgenstern, 1944), at the same time psychology and economics have a long tradition in the development of stochastic choice models (e.g., Luce, 1959; Train, 2003). Choice rules are important because empirical research has shown that choice behavior is probabilistic in that people do not always make the same choice even in nearly identical choice situations (Hey, 2001; Mosteller & Nogee, 1951). When explicitly modeling this variability, researchers have to make an assumption on where the random component affects the decision process: Loomes, Moffatt, and Sugden (2002) distinguished randomness in assigning utility to options, randomness in comparing different options with each other, and randomness in the implementation of a decision. The first approach is best characterized by random utility models (e.g., Becker, DeGroot, & Marschak, 1963; Train, 2003). In these models, utility itself is a random variable. This can be motivated by the assumption that people estimate utilities with respect to different aspects of an option. The second approach assumes fixed utility but a choice function that introduces randomness in the comparison stage (e.g., Becker et al., 1963; Bridle, 1990; Luce, 1959). Finally, a trembling hand error—adding the probability of choosing the inferior of two options independent of the difference in expected utilities of the options—is an example of randomness in choice implementation (Harless & Camerer, 1994; Selten, 1975).

Which decision process model is most accurate is an empirical question and the subject of active debate. Blavatskyy and Pogrebna (2010), for example, examined different stochastic choice models and concluded that a Fechner model with heteroscedastic and truncated random errors fit data better than a Fechner model with homoscedastic error components. Using a probit or a trembling hand model is definitely limiting as, for example, they fail to account for context effects (Wilcox, 2015). Context effects occur if choice behavior depends on the choice set presented, and Rieskamp, Busemeyer, and Mellers (2006) summarized empirical evidence for it. Our approach here, however, was not meant to find the theory that describes the data best. Rather, we used two relatively simple stochastic choice models (probit and trembling hand) to measure the relative influence of systematic changes in preference and choice consistency, respectively. Yet, the examination of the effect of a reduction in cognitive capacities on more complex choice models might also be warranted.

Reduced Cognitive Capacities in the Real World

In general, the dual-task design implemented in our work is meant to capture a ubiquitous phenomenon in our daily life, namely, decision making under reduced cognitive capacities. Cognitive capacities can be limited for many reasons, such as multitasking, stress, sleep deprivation, alcohol consumption, or a lack of motivation (e.g., Anderson & Dickinson, 2010; Morgado et al., 2015). Because people have to make decisions under such circumstances quite frequently, it is an important question how behavior might differ compared to behavior in situations with full cognitive capacities.

What would one expect to happen when making decisions under reduced cognitive capacities in real life? One prominent answer comes from the nudge program (Thaler & Sunstein, 2008), which suggests that people do not always make decisions in line with their (long-term) goals and sometimes need assistance to improve or "debias" their decisions. Taking our results into account, however, there is no indication of a need for debiasing preferential shifts under reduced cognitive capacities because deviations from true preference can equally likely go in either direction (e.g., more or less risk taking). Rather, one would expect participants to be less predictable in their choice behavior under reduced cognitive capacities. This might be bad when people make a decision once and most likely stick to this decision for a long time as, for example, in retirement savings decisions. As a result, many people do not save according to their true preferences for future consumption (Skinner, 2007). On the other hand, looking at repeated small decisions such as in grocery shopping, deviations from true preferences might cancel out after many bargains. In addition, deviating from a previous choice might be advantageous in that it boosts learning in changing environments. Hence, our results explain why people sometimes make inconsistent decisions and we predict seeing these inconsistencies more often when cognitive capacities are reduced.

The present study was designed to unify research on decision making in preferential and economic choice with recent work on the effect of cognitive capacity limitations. The mathematical modeling approach that we used can be applied, in principle, to all domains of preferential choice as long as preferences can be mathematically specified. Furthermore, the models allow exploration of the cognitive similarities and differences between manipulations such as cognitive load, ego depletion, time pressure, and sleep deprivation, among others. Thus, research in the field of reduced cognitive capacities can profit from the mathematical approach presented here, that is, an explicit formulation of both the underlying preferential choice model and the stochastic choice rule.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82,* 463.

Anderson, C., & Dickinson, D. L. (2010). Bargaining and trust: The effects of 36-h total sleep deprivation on socially interactive decisions. *Journal of Sleep Research, 19,* 54–63.

Andersson, O., Tyran, J.-R., Wengström, E., & Holm, H. J. (2013). *Risk aversion relates to cognitive ability: Fact or fiction?* IFN Working Paper No. 964.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16,* 351–355.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science, 8,* 41–55.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9,* 226–232.

Benjamin, D. J., Brown, S. A., & Shapiro, J. M. (2013). Who is 'behavioral'? cognitive ability and anomalous preferences. *Journal of the European Economic Association, 11,* 1231–1255.

Birnbaum, M. H., & Bahra, J. P. (2012). Separating response variability from structural inconsistency to test models of risky decision making. *Judgment and Decision Making, 7,* 402.

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25,* 963–986.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90,* 166–193.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior, 10,* 95–121.

Bridle, J. S. (1990). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems* (pp. 211–217).

Brosnan, S., & de Waal, F. (2014). Evolution of responses to (un) fairness. *Science*. Advance online publication. http://dx.doi.org/10.1126/science.1251776

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 7745–7750.

Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives, 9,* 209–219.

Cappelletti, D., Güth, W., & Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology, 32,* 940–950.

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology, 5*.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization, 87,* 43–51.

Cohen, J. D., Forman, S. D., Braver, T. S., Casey, B., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of the prefrontal cortex in a nonspatial working memory task with functional MRI. *Human Brain Mapping, 1,* 293–304.

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review, 78,* 97–119.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology, 59,* 1070–1100.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review, 100,* 1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9,* 522–550.

Ebert, J. E. (2001). The role of cognitive resources in the valuation of near and far future events. *Acta Psychologica, 108,* 155–171.

Ebert, J. E., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management science, 53,* 1423–1438.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, August,* 817–868.

Fisher, I. (1930). *The theory of interest*. New York, NY: MacMillan.

Fleischhut, N., Artinger, F., Olschweski, S., Volz, K. G., & Hertwig, R. (2014). Sampling of social information: Decisions from experience in bargaining. In

P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1048–1053). Austin, TX: Cognitive Science Society.

Franco-Watkins, A. M., Pashler, H., & Rickard, T. C. (2006). Does working memory load lead to greater impulsivity? commentary on Hinson, Jameson, and Whitney (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 443–447.

Franco-Watkins, A. M., Rickard, T. C., & Pashler, H. (2010). Taxing executive processes does not necessarily increase impulsive decision making. *Experimental Psychology, 57,* 193–201.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40,* 351–401.

Freeman, N., & Muraven, M. (2010). Self-control depletion leads to increased risk taking. *Social Psychological and Personality Science, 1,* 175–181.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: Chapman & Hall/CRC.

Gevins, A., & Cutillo, B. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and Clinical Neurophysiology, 87,* 128–143.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130,* 769.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3,* 367–388.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136,* 495.

Halali, E., Bereby-Meyer, Y., & Meiran, N. (2014). Between self-interest and reciprocity: The social bright side of self-control failure. *Journal of Experimental Psychology: General, 143,* 745.

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica: Journal of the Econometric Society, 62,* 1251–1289.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review, 39,* 633–640.

Hey, J. D. (2001). Does repetition improve consistency? *Experimental Economics, 4,* 5–54.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society, 62,* 1291–1326.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 298.

Hornik, K., Leisch, F., & Zeileis, A. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of DSC* (Vol. 2, p. 1).

Johnson, E. J. (2008). Man, my brain is tired: Linking depletion and cognitive effort in choice. *Journal of Consumer Psychology, 18,* 14–16.

Joireman, J., Balliet, D., Sprott, D., Spangenberg, E., & Schultz, J. (2008). Consideration of future consequences, ego-depletion, and self-control: Support for distinguishing between cfc-immediate and cfc-future subscales. *Personality and Individual Differences, 45,* 15–21.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review, 76,* 728–741.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615.

Knafo, A., Zahn-Waxler, C., Van Hulle, C., Robinson, J. L., & Rhee, S. H. (2008). The developmental origins of a disposition toward empathy: Genetic and environmental contributions. *Emotion, 8,* 737–752.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. New York, NY: Academic Press.

Law, A. S., Logie, R. H., & Pearson, D. G. (2006). The impact of secondary tasks on multitasking in a virtual environment. *Acta Psychologica, 122,* 27–44.

Levy, H., & Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review,* 308–317.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty, 24,* 103–130.

Luce, D. R. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science, 19,* 255–260.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience, 27,* 5796–5804.

Meiser, T., Klauer, K. C., & Naumer, B. (2001). Propositional reasoning and working memory: The role of prior training and pragmatic content. *Acta Psychologica, 106,* 303–327.

Miller, K., Price, C., Okun, M., Montijo, H., & Bowers, D. (2009). Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology, 24,* 711–717.

Morgado, P., Sousa, N., & Cerqueira, J. (2015). The impact of stress in decision making in the context of uncertainty. *Journal of Neuroscience Research, 93,* 839–847.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy, 59,* 371.

O'Donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making, 13,* 233.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 534.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Peirce, J. W. (2007). Psychopy—Psychophysics software in python. *Journal of neuroscience methods, 162,* 8–13.

Perlstein, W. M., Dixit, N. K., Carter, C. S., Noll, D. C., & Cohen, J. D. (2003). Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biological Psychiatry, 53,* 25–38.

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PloS ONE, 7,* e47225.

Phillips, L., Gilhooly, K., Logie, R., Della Sala, S., & Wynn, V. (2003). Age, working memory, and the Tower of London task. *European Journal of Cognitive Psychology, 15,* 291–312.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society, 32,* 122–136.

R Core Team. (2016). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: Author. Retrieved from https://www.R-project.org/

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1446.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature, 44,* 631–661.

RStudio Team. (2015). *R* [Computer software manual]. Boston, MA: Author. Retrieved from http://www.rstudio.com/

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies, 4,* 155–161.

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology, 41,* 77–87.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4,* 25–55.

Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R., . . . Gray, J. R. (2008). Individual differences in delay discounting: Relation to intelligence, working memory, and anterior prefrontal cortex. *Psychological Science, 19,* 904–911.

Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research, 26,* 278–292.

Skinner, J. (2007). Are you sure you're saving enough for retirement? *The Journal of Economic Perspectives, 21,* 59–80.

Stewart, N., Reimers, S., & Harris, A. J. (2014). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61,* 687–705.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thurstone, L. L. (1927). Psychophysical analysis. *The American Journal of Psychology, 38,* 368–389.

Train, K. (2003). *Discrete choice with simulation*. New York, NY: Cambridge University Press.

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

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505.

Vehtari, A., Gelman, A., & Gabry, J. (2016a). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models* (R package version 1.1.0). Retrieved from https://CRAN.R-project.org/package=loo

Vehtari, A., Gelman, A., & Gabry, J. (2016b). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27,* 1413–1432.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wilcox, N. (2015). *Error and generalization in discrete choice under risk*. Working Paper.

Worthy, D. A., Otto, A. R., & Maddox, W. T. (2012). Working-memory load and temporal myopia in dynamic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1640–1658.

Appendix A

Model Recovery to Distinguish Choice Inconsistency From Preference

We conducted a simulation to show the general ability of our model to distinguish between shifts in preferences and shifts in choice inconsistencies. Therefore, we used the 400 trials we created for the first experiment and extracted the average parameter values for risk preference ($M_{Pref} = -0.05$) and choice sensitvitiy ($M_{Sensitvitiy} = -1.68$) from the experimental data assuming a linear utility and a probit choice model as explained in the main text. With these data we created two times 80 choices for 40 synthetic participants in two conditions. In the control condition we created choices based on the average parameter values. For the load condition we changed either the preference or the choice inconsistency parameter in the magnitude of one standard deviation ($SD_{Pref} = 0.08$, $SD_{Sensitivity} = 0.28$) to simulated shifts in preferences or choice inconsistencies. The standard deviations were taken from the empirical distributions of individual parameter estimates of the first experiment.

To fit the simulated data we used the hierarchical Bayesian model as described in the Method section of Experiment 1, using a linear utility and a probit choice function. In total we ran the simulation and the following model recovery analysis 100 times. We implemented two types of choice sets with the aim of demonstrating the robustness of the approach under conditions both where the choice proportions under control were around 50% and where they were biased away from 50%. For the first choice set, we implemented a set of choices across the whole spectrum of expected value (EV) differences between the safer and the riskier option (see Experiment 1 Method section). The upper row in Table A1, table in the Appendix A shows the results. First, we checked whether choice proportions were different in the control compared with the load condition by means of a paired *t*-test. Choice proportions differed in only 11 out of 100 simulations when choice consistency was manipulated, but did so in all cases when risk preferences were changed. The 11 significant choice proportion differences with simulated sensitivity shifts resulted from unlikely choice proportions relatively far away from 50% in the control condition, which were then dragged towards 50% due to a higher simulated noise in the load condition. For the parameter recovery, we applied the 95%–HDI approach and classified a recovered parameter shift whenever the 0 were excluded from this interval in the posterior distribution of parameter differences. We conclude that the difference in the true parameter can be recovered perfectly and that in less than 5% of the cases the unchanged parameter was estimated to be different.

In a second step, we created a choice set where choice proportions were different from 50% in the control condition by using choice situations where the riskier of the two options had a much lower expected value than the safer option. This resulted in a low choice proportion of the risky option of 36%. Here, we observed that both a shift in choice inconsistency as well as a shift in preferences change choice proportions in a systematic way. Significant differences in choice proportions were always a shift from below 50% (mean control 36%) towards 50%. The mean choice proportion in the load condition was 42% for a simulated increase in choice inconsistency and 49% for a simulated shift in risk preference. This is the case because more inconsistencies result in a choice proportion in the control condition closer to a chance level of 50%. Also, with this choice set, the true source of difference can be distinguished almost perfectly. The risk of concluding from our modeling framework that the actually unchanged parameter shows a significant difference is small (below 10%).

Table A1
Simulation Risky Gambles: Choice Proportion Differences and Recovered Parameter Differences in 100 Simulations

	Simulate sensitivity shift			Simulate preference shift		
Stimuli	Difference choice prop.	Recovered risk shift	Recovered sensitivity shift	Difference choice prop.	Recovered risk shift	Recovered sensitivity shift
Unbiased (50% in control)	11%	3%	100%	100%	100%	2%
Biased (36% in control)	94%	1%	99%	100%	100%	7%

Note. Data were simulated and recovered by a model with linear utility and a probit choice function. For additional model specifications see introduction and the Method section of Experiment 1. Significant differences in choice proportions were assessed with a paired *t*-test at the 1% significance level. Significant differences in model parameters between control and load condition were inferred according to the 95%–HDI criterion.

(*Appendices continue*)

Appendix B

Correlations of Model Parameters With Behavioral Measures

Table B1

Experiment 1 Risky Gambles: Correlations of Model Parameters With Behavioral Measures

Variable/Parameter	*N*-back	Ospan	RT_{load}—RT_{control}	Preference	$\delta_{\text{preference}}$	Sensitivity	$\delta_{\text{sensitivity}}$
N-back	—						
Ospan	.19	—					
RT_{load}—RT_{control}	−.17	.08	—				
Preference	.20	.23	.09	—			
$\delta_{\text{preference}}$	.12	.03	.12	.06	—		
Sensitivity	−.07	−.21	−.43**	−.05	.09	—	
$\delta_{\text{sensitivity}}$	−.14	.15	.06	.21	−.43**	.00	—

Note. Model parameters are taken from a power utility with probit error model as described in the main text.
** $p < .01$.

Table B2

Experiment 2 Temporal Discounting: Correlation of Model Parameters With Behavioral Measures

Variable/Parameter	*N*-back	Ospan	RT_{load}—RT_{control}	Preference	$\delta_{\text{preference}}$	Sensitivity	$\delta_{\text{sensitivity}}$
N-back	—						
Ospan	−.07	—					
RT_{load}—RT_{control}	−.73***	.03	—				
Preference	.26	.14	−.25	—			
$\delta_{\text{preference}}$	−.20	.07	.13	.07	—		
Sensitivity	−.29	−.21	.31*	−.63***	−.02	—	
$\delta_{\text{sensitivity}}$	.05	−.09	.08	.06	−.52***	.03	—

Note. Model parameters are taken from a one-parameter hyperbolic discounting function and a normally distributed error around the discounted outcome as outlined in the main text.
* $p < .05$. *** $p < .001$.

Table B3

Experiment 3 Miniultimatum Game: Correlation of Model Parameters With Behavioral Measures

Variable/Parameter	*N*-back	Ospan	RT_{load}—RT_{control}	Preference	$\delta_{\text{preference}}$	Sensitivity	$\delta_{\text{sensitivity}}$
N-back	—						
Ospan	.01	—					
RT_{load}—RT_{control}	−.37**	.02	—				
Preference	.10	−.02	−.14	—			
$\delta_{\text{preference}}$	.11	−.22	.12	−.01	—		
Sensitivity	.00	−.15	−.08	.30*	.06	—	
$\delta_{\text{sensitivity}}$	−.13	−.15	−.05	−.09	−.13	.12	—

Note. The model parameters are taken from the first-order inequity aversion model of Fehr and Schmidt (1999) combined with a probit error model as described in the main text.
* $p < .05$. ** $p < .01$.

Received October 27, 2016
Revision received November 27, 2017
Accepted December 2, 2017 ■