TI 2021-012/I
Tinbergen Institute Discussion Paper

The Informational Affective Tie Mechanism: On the Role of Uncertainty, Context, and Attention in Caring

Frans van Winden[1]

---

[1] Emeritus Professor Amsterdam School of Economics - CREED and Amsterdam Brain & Cognition University of Amsterdam

The Informational Affective Tie Mechanism

On the Role of Uncertainty, Context, and Attention in Caring

Frans van Winden

Emeritus Professor
Amsterdam School of Economics - CREED
and
Amsterdam Brain & Cognition
University of Amsterdam
f.a.a.m.vanwinden@uva.nl

ABSTRACT

This paper proposes and applies a formal theoretical model of an automatic (prosocial and antisocial) caring mechanism: the informational affective tie mechanism (*i*ATM) model. Novel in the formalization is the factoring in of information extraction concerning the behavioral type of agents interacted with, jointly with the contexts of these interactions and the attention they attract. Empirical support comes from five very different data sources: experimental findings, econometric results, model-based brain scanning (fMRI) findings, additional neurobiological evidence, and translational and evolutionary evidence. Applications address: the impact of time pressure and cognitive load; the endogeneity of different behavioral response patterns (like tit-for-tat); social preference drift and tipping points in collective action; why behavioral survey questions can be problematic; spread of caring through affective networks, an uncertainty-based link between social-, risk- and time-preferences; happiness and identity; and, the neglected political economic role of communities (next to centralized authorities and markets). The endogeneity of caring preferences sharply contrasts with the standard assumption in economic theory of stable (mostly selfish) preferences. Moreover, the provision of a neurobiological underpinning moves the *i*ATM model away from the standard as-if approach towards an as-is approach. Although the focus is on humans, some attention will be paid also to the model's relevance for studying other species.

January 2021

1 Introduction

The importance of allowing for affective social relationships and prosocial behavior, defined as the voluntary provision of a benefit to another agent which is costly to the provider, has become widely acknowledged in economics and other social sciences (see, e.g., Akerlof 1983, Coleman 1984, Granovetter 1985, Uzzi 1999, Van Lange 1999, Fehr and Gaechter 2002, Sobel 2005, Bénabou and Tirole 2006, Leider et al. 2009, Cooper and Kagel 2016). Prosocial behavior is also increasingly reported and studied in the life sciences, not only regarding other primates like bonobos but also concerning many other species across the evolutionary ladder, which may involve genetically unrelated strangers (see, e.g., Silk 2002, de Waal 2008, Schino and Aureli 2009, Massen et al. 2010, Seyfarth and Cheney 2012, Cronin 2012). Well-known examples referred to concern mammalians, such as rodents like rats and voles (Ben-Ami Bartal et al. 2014, Insel and Young 2001). However, more recent evidence even includes birds (e.g., parrots and bats), plants, and bacteria (Brucks and von Bayern 2020, Carter et al. 2020, Kiers et al. 2011, West et al. 2002, Whiteley et al. 2017). Note that species are included without higher-order cognition enabling calculated reciprocity (planned strategic behavior). As a consequence, there are now many overlapping questions in the behavioral and life sciences concerning the (proximate) driving factors of the observed prosocial behavior and the modeling thereof. Because of the continuity in evolution – where new is typically building on but not replacing old – a similar type of underlying mechanism may be conjectured, which might be related to the behavioral uncertainty regarding interaction counterparts shared by all organisms. According to Damasio (2018, p55): “The principle is always the same: organisms give up something in exchange for something that other organisms can offer them; in the long run, this will make their lives more efficient and survival more likely. What bacteria, or nucleated cells, or tissues, or organs give up, in general, is independence; what they get in return is access to the “commons,” the goods that come from a cooperative arrangement (…).” With the proviso that the benefit of the exchange need not necessarily go to the organism giving up something (think of prosocial behavior among kin), the reference to greater efficiency suggests that, in a “commons” context, an agent (organism) would somehow and to some extent – based on information extracted from the experienced behavior of another agent – internalize the external effect of its own behavior on the agent interacted with. This internalization, pointing at an intrinsic motivation to choose an action affecting the other agent’s well-being, will be labeled “*caring*” in this paper. While the above refers to prosocial behavior (positive caring), it should be immediately added that antisocial behavior (negative caring), that is, the costly infliction of a detriment on another agent (like a sanction), is also increasingly observed across the same evolutionary ladder (Fehr and Gaechter 2000, Wiegman 2019, West et al. 2002, Kiers et al. 2011). This is consistent with the view that both prosocial and antisocial behavior are important for survival and fitness – for instance, to avoid the exploitation of prosocial behavior –, and should therefore jointly be allowed for in a behavioral model of organisms. In very general terms, and more formally, such a behavioral caring model may be represented as follows. Denoting an action by agent i (j), with an external effect on j (i), by a_i (a_j), the probability with which a_i is chosen by $p(a_i)$, and i’s care for j by $care_{ij}$: $p(a_i) = f(a_i; care_{ij})$ with $care_{ij} = g(a_j)$, where f and g stand for functional relationships. Key is the mediating role of i’s care for j in f, which changes the nature of the choice valuation function and would represent the above referred to underlying mechanism for prosocial (or antisocial) behavior. The precise nature of such a caring mechanism would differ between species, for instance, based on relatively simple chemical (hormonal) responses in bacteria to more complicated chemico-electric (hormonal and neural) responses in animals like humans.

These findings and considerations, and substantial further evidence referred to below, inspire the following *basic hypothesis* underlying the formal theoretical and empirical analysis of this paper:

agents facing environmental uncertainty, where other agents may turn out to be benefactors or malefactors, will automatically develop a positive or negative (emotive) action tendency regarding an agent interacted with, based on the information regarding the nature of that agent extracted from its behavior; this action tendency reflects an intrinsic motivation to seek the other's proximity, or to keep a distance, and to provide benefits or detriments, that is, to care for that agent.

In humans, mostly focused on below, this assumedly primordial caring mechanism perhaps most clearly manifests itself in the attachment of a newborn to its primary caregiver (which need not be genetically related). Human attachment theory assumes an innate and automatic attachment system, involving both infant and caregiver, which motivates an infant to seek proximity to its caregivers in times of need or the presence of threat, which is crucial for its survival (Vrticka and Vuilleumier 2012; see further below). Clearly, in humans (seeking) help may be induced as well by the higher-order cognitive (strategic) reasoning processes that are typically focused on in formal economic models and game theory. This is in line with Kahneman (2011) who distinguishes between two mental systems. First, a non-strategic affective mechanism rooted in evolutionary older (limbic) parts of the brain, characterized by automaticity, speed, and impulsivity (labeled System I). Second, a higher-order cognitive reasoning mechanism that is particularly rooted in the prefrontal neocortex of the brain, characterized by slower, more deliberate and effortful processes (labeled System II). Importantly, the relative impact of System II ("cognition") versus the evolutionary older System I ("emotion") is related to self-control and relies on the presence of sufficient mental resources; there is a primacy of affect such that System I may dominate decision-making (Zajonc 1984, LeDoux 1998, Kahneman 2011). This makes the affective caring mechanism proposed in this paper of significance for humans also. A similar kind of distinction between cognitive and affective mechanisms holds for empathy that plays an important role in human caring, as further discussed below. While cognitive empathy refers to understanding what something means for someone else (also called mentalizing or theory of mind), emotional or affective empathy relates to the sharing of emotions, with each type of empathy having its own characteristic neural circuitry (Shamay-Tsoory etal. 2009).

This paper proposes a formal theoretical model of this caring system, dubbed the informational affective tie mechanism model or *i*ATM model, for short. Novel in this formalization, which is inspired by the social ties model of van Dijk and van Winden (1992, 1997), is the explicit factoring in of information extraction concerning the behavioral type of agents interacted with, jointly with the contexts of these interactions and the attention they attract. Substantial empirical support from the following five data sources will be provided: experimental findings, model-based econometric results (parameter estimates), model-based brain scanning (fMRI) results, further neurobiological evidence, and evolutionary (translational) evidence. This is followed by a number of applications dealing with: the impact of time pressure and cognitive load; the endogeneity of different behavioral response patterns (like tit-for-tat); social preference drift and tipping points in collective action; why behavioral survey questions can be problematic; spread of caring through affective networks, an uncertainty based link between social-, risk- and time-preferences; happiness and identity; and, the neglected political economic role of communities (next to centralized authorities and markets). The endogeneity of preferences related to caring sharply contrasts with standard formal economic theory where stable preferences are assumed. However, it fits an emerging relationship science, the goal of which is to understand relationship dynamics and the antecedents and consequences of these (Reiss et al. 2000). The provision of a neurobiological underpinning, moreover, moves the *i*ATM model away from the standard economic "as if" approach towards an "as is" (because) approach (Glimcher 2011). It allows, furthermore, for the development of both prosocial as well as antisocial relationships, together with the associated formation of (positive, negative or mixed) affective networks, issues that are typically studied seperately in the literature. Finally, the model is flexible and can be straightforwardly expanded to incorporate strategic reasoning (such as calculated reciprocity). Although the focus is on humans, some attention will be paid also to the model's relevance for studying other species.

The organization of this paper is further as follows. Section 2 presents the formalization of the informational affective tie mechanism, the *i*ATM model. Section 3 goes into the five sources of evidence regarding the model and discusses some implications. Applications, subsumed under ten items, are provided in Section 4, while Section 5 concludes.

2 The Informational Affective Tie Mechanism

This section presents and formalizes the informational affective tie mechanism (*i*ATM) and fits it into a decision model. The next two sections provide empirical support and a number of applications.

The iATM Model

The informational affective tie mechanism basically consists of three modules. The first module concerns a *friend-foe appraisal*, involving an experiential assessment of the true reward contingency (type) of an agent interacted with. A "friend" is associated with a predicted positive change in welfare or utility (these two concepts are used interchangeably), while a negative change is associated with a "foe". Or, put differently, friends (foes) are expected to be positively (negatively) caring about one's welfare. The second module deals with *affective tie formation*, formalizing the affective bond with the type of agent interacted with, given the context attended to. The third module, finally, regards a spillover or generalization effect, coined a *generalized tie value* (*GTV*, for short). GTV formalizes the affective tie value concerning a *generalized other*, that is, an agent assessed as novel (like an anonymous randomly selected agent).

Module 1: Friend-Foe Appraisal

This module assumes that the appraisal of the (friend or foe) type of an agent involves an optimal experiential assessment, based on the interaction with that agent. Let τ_{ijt} denote the true reward contingency for agent *i* of meeting agent *j* at time *t*, labeled *j*'s *type*. (Note that a reward can be negative.) Types are allowed to range from extreme foe ($-\infty$) to extreme friend ($+\infty$), that is: $-\infty < \tau_{ijt} < +\infty$.

Behavior of *j* at *t*, determining the actual reward to *i*, generates an *impulse* I_{ijt}, experienced by *i* as a signal of *j*'s type. An impulse is assumed to be stochastically related to *j*'s type: $I_{ijt} = \tau_{ijt} + \varepsilon_t$, where ε_t is taken to be an independent Gaussian distributed noise term, with zero mean and variance σ_ε^2 reflecting *behavioral uncertainty* (unaccounted for factors influencing *j*'s behavior, whatever its type).

The experienced impulse I_{ijt} concerns a normalization, based on the action a_{jt} taken by *j*:

$$(1)\quad I_{ijt} = \left(a_{jt} - a_{ijt}^{ref}\right) / \left(a_{ijt}^{eff} - a_{ijt}^{ref}\right)$$

where a_{ijt}^{ref} denotes a *reference action*, expected from a *j* who is neither friend nor foe (that is, a non-caring type), and a_{ijt}^{eff} stands for an *efficient action*, that is, a cooperative action by *j* maximizing the joint welfare of *i* and *j* (thereby internalizing any external effect of its behavior on *i*'s welfare); for an interesting analysis $a_{ijt}^{eff} \neq a_{ijt}^{ref}$ is assumed to hold. Note that $I_{ijt} = 0$ if *j* takes the reference action, while $I_{ijt} = 1$ if *j* takes an efficient action (see further Module 2).

Let agent *i*'s *prior* appraisal of τ_{ijt} be Gaussian distributed with mean α_{ijt} and variance σ_{ijt}^2 reflecting *type uncertainty*. Following an impulse, agent *i* optimally (Bayesian) updates its prior to a *posterior* appraisal α_{ijt+1}. It can be proved (see Appendix) that this posterior appraisal will be normally distributed, with mean:

$$(2)\ \ \alpha_{ijt+1} = \alpha_{ijt} + \delta_{ijt}(I_{ijt} - \alpha_{ijt}) = (1 - \delta_{ijt})\alpha_{ijt} + \delta_{ijt}I_{ijt}$$

and variance:

$$(3)\ \ \sigma_{ijt+1}^2 = (1 - \delta_{ijt})\sigma_{ijt}^2$$

where:

$$(4)\ \ \delta_{ijt} = \sigma_{ijt}^2 \,/\, (\sigma_{ijt}^2 + \sigma_\varepsilon^2) = 1/(1 + \sigma_\varepsilon^2/\sigma_{ijt}^2).$$

Note from eq. (2) that repeatedly cooperative behavior by *j* (that is, $I_{ij} = 1$) would move the weight attached by *i* to the utility of *j* (α_{ij}) towards 1, making *i* in turn more likely to become cooperative towards *j*, which has a reinforcing effect on *j*'s behavior. Note from eq. (4), furthermore, that the updating factor δ_{ijt} – the learning rate – only depends on the ratio of behavioral uncertainty to type uncertainty ($\sigma_\varepsilon^2/\sigma_{ijt}^2$); this ratio increases with more interaction experiences (impulses), as they diminish the type uncertainty (see eq. (3), with a smaller impact of further impulses as consequence (eqs. (2) and (4)).

Module 2: Affective Tie Formation, Context, and Attention

The *key assumption* of this module is that an agent's type appraisal (α) generates a weight attached to the welfare or utility of that agent, which reflects an interaction-experience based *affective tie* inducing an intrinsic motivation to care for that agent. By implication, preferences become endogenous, for dependent on social interaction experiences.

The assessment of an agent's type may be more or less reliable, however, with *reliability* being a (negative) function of the variance σ_{ijt}^2, denoted by $f(\sigma_{ijt}^2)$. Because unreliability can be seen as a kind of risk – namely, *type risk* – that agents may or may not like, a more general expression of an *affective tie*, denoted by $\bar{\alpha}_{ijt}$, would be: $\bar{\alpha}_{ijt} = f(\sigma_{ijt}^2)\alpha_{ijt}$. In case of type-risk neutrality, $f(\cdot)$ would be a constant function with the constant being equal to one (thus, $\bar{\alpha}_{ijt} = \alpha_{ijt}$), while type-risk aversion would imply a negative first-order derivative, denoted by $f'(\cdot) < 0$ (with $\bar{\alpha}_{ijt} \leq \alpha_{ijt}$), and type-risk seeking a positive first-order derivative, denoted by $f'(\cdot) > 0$ (with $\bar{\alpha}_{ijt} \geq \alpha_{ijt}$). For illustration, the following simple specification: $f(\sigma_{ijt}^2) = e^{-\sigma_{ijt}^2}$ could hold for risk-aversion, and $f(\sigma_{ijt}^2) = e^{\sigma_{ijt}^2}$ for risk-seeking. Note that, whatever the risk attitude, $\bar{\alpha}_{ijt} = \alpha_{ijt}$ if the appraisal of *j*'s type is deemed to be fully reliable ($\sigma_{ijt}^2 = 0$). Furthermore, in case of type-risk aversion, the tie would get closer to 0 the larger σ_{ijt}^2, that is, the more unreliable the appraisal of *j*'s type becomes.

Because information extraction resources are limited, the extent to which certain experiences will attract (un)conscious attention in the decision-making process may vary. This will be dealt with by applying an *attentional weight* ($0 \leq \gamma \leq 1$) to an interaction context, reflecting its memory association strength (more on this below).

Now, first assume that *i* only interacts with *j* within one particular context $\mathbb{C}$, with attentional weight $\gamma_{i\mathbb{C}}$. Letting $\bar{\alpha}_{ij}$ denote the affective tie with *j*, and U_j the utility of *j* (as perceived by *i*), the following *extended utility* (V_i) representation is assumed to hold:

(5) $V_{it} = U_{it} + \gamma_{i\mathbb{C}t}\bar{\alpha}_{ijt}U_{jt}$

Module 3: Generalized Tie Value

This third, and final, module addresses what happens if, within the same context, *i* subsequently meets agent *k* in that context (a generalized other). In that case, *i* is assumed to generalize its type appraisal based on its interaction experience so far. Specifically, the prior mean appraisal of *k*'s type, α_{ikt}, is assumed to equal *i*'s present appraisal of *j*, α_{ijt}; thus, $\alpha_{ikt} = \alpha_{ijt}$. Because of the lack of experience with the new agent, and for simplicity, the prior variance is taken to equal a fixed initial variance denoted by σ_0^2. Consequently, *i* would start the interaction with an affective tie regarding *k* equal to: $\bar{\alpha}_{ikt} = f(\sigma_0^2)\alpha_{ijt}$. Note that even with no past or future interaction with *k*, *i* would still to some extent care for *k* in case of a non-zero tie value with *j*. Because of this spillover or generalization effect, we will call this tie value a *generalized tie value* (*GTV*). In this case:

(6) $GTV_{it} = \gamma_{i\mathbb{C}t}\bar{\alpha}_{ikt} = \gamma_{i\mathbb{C}t}f(\sigma_0^2)\alpha_{ijt}$

Now, let C denote the set of agents interacted with in context $\mathbb{C}$, with $c\epsilon C$ as characteristic element, and cardinality $|C|$, then *i*'s extended utility can be written as:

(5a) $V_{it} = U_{it} + \gamma_{i\mathbb{C}t}\sum_{c\epsilon C}\bar{\alpha}_{ict}\,U_{ct}$

while the GTV regarding context $\mathbb{C}$ becomes:

(6a) $GTV_{it} = \gamma_{i\mathbb{C}t}\sum_{c\epsilon C}f(\sigma_0^2)\alpha_{ict}\;/\;|C|$

Denoting the utility of a generalized other by U_g, and the extended utility in a novel interaction with a generalized other by V_i^g, renders:

(7) $V_{it}^g = V_{it} + GTV_{it}{\cdot}U_{gt}$

Incidentally, note that our focus thus far (and below) is on individual-specific ties, requiring that agents can recognize each other and become specific agents to each other. If unrecognizable (i.e., perceived as identical), the other agent, whoever it is, would be like a single agent interacted with. In that case, the same specification is assumed to hold as for a specific agent (like eq. (2)), even though the actions may stem from different agents.

Note, furthermore, that agents may take into account an interaction partner's extended utility *V*, instead of its direct utility *U*. Empathic skills are obviously relevant here.

Contexts

Any interaction takes place within a certain context, and together they make up an interaction episode that may be more or less easily remembered dependent on the nature of the context, its timing, and the hedonic value or experienced utility (Kahneman et al. 1997) of the experienced interaction (determining its salience and emotional imprint). Among the defining factors of a *context* are likely to be: the type of game that is played (with an important horizontal competition-cooperation dimension, and a vertical hierarchy or dominance dimension), the type(s) of agent(s) involved (where uncertainty may be related to nature, culture, and existing ties with the protagonist), and any other uncertainty influencing behavior apart from type uncertainty.

Now, if interaction is going to take place within a novel context, uncertainty about agent types and their reliability is likely to be affected, dependent on the similarity of the new context with earlier experienced contexts. Assuming that similarity, like timing and hedonic value for that matter, is captured by the attentional weight (association strength) of a context, the next equations generalize the above expressions for extended utility and the generalized tie value. Let $\mathfrak{C}$ stand for the set of relevant contexts: $\mathfrak{C} = \{\mathbb{C}1, \mathbb{C}2, \dots, \mathbb{C}\mathrm{N}\}$, with characteristic element $\mathbb{C}$. Furthermore, again, let the set of agents in $\mathbb{C}$ be denoted by C, with characteristic element *c* and cardinality $|C|$. Then, extended utility can be written as:

(5b) $V_{it} = U_{it} + \gamma_{i\mathbb{C}1t} \sum_{c1 \epsilon C1} \bar{\alpha}_{ic1t}\, U_{c1t} + \gamma_{i\mathbb{C}2t} \sum_{c2 \epsilon C2} \bar{\alpha}_{ic2t}\, U_{c2t} + \cdots + \gamma_{i\mathbb{C}Nt} \sum_{cN \epsilon CN} \bar{\alpha}_{icNt}\, U_{cNt}$

with: $0 \leq \sum_{\mathbb{C} \epsilon \mathfrak{C}} \gamma_{i\mathbb{C}t} \leq 1$, while the generalized tie value becomes:

(6b) $GTV_{it} = \sum_{\mathbb{C} \epsilon \mathfrak{C}} \gamma_{i\mathbb{C}t} \sum_{c \epsilon C} f(\sigma_0^2) \alpha_{ict} \;/\; |C|$

Volatility

Finally, we want to account for *volatility*, more specifically, the possibility of repeated random shocks to the true reward that can be expected from a counterpart, for example, because of a changing environment due to migration (the next section also refers to the issue of surprises, that is, unexpected impulses). Assume that $\tau_{ijt} = \tau_{ijt-1} + \eta_t$, where η_t stands for an independent Gaussian noise term, with zero mean and variance σ_{shock}^2. In that case, $(\sigma_{ijt}^2 + \sigma_{shock}^2)$ replaces σ_{ijt}^2 in the posterior variance σ_{ijt+1}^2 (eq. (3)) and the learning rate δ_{ijt} (eq. (4)):

(3a) $\sigma_{ijt+1}^2 = \left(1 - \delta_{ijt}\right)\left(\sigma_{ijt}^2 + \sigma_{shock}^2\right)$

(4a) $\delta_{ijt} = \frac{\sigma_{ijt}^2 + \sigma_{shock}^2}{\sigma_{ijt}^2 + \sigma_{shock}^2 + \sigma_{\varepsilon}^2}$

Furthermore, $f\left(\sigma_0^2 + \sigma_{shock}^2\right)$ has to be substituted for $f(\sigma_0^2)$ in the GTV expression (eq. (6b)):

(6c) $GTV_{it} = \sum_{\mathbb{C} \epsilon \mathfrak{C}} \gamma_{i\mathbb{C}t} \sum_{c \epsilon C} f\left(\sigma_0^2 + \sigma_{shock}^2\right) \alpha_{ict} \;/\; |C|$

Action selection

Incorporating the *i*ATM into a decision model will of course generally affect behavior. In empirical decision models often a soft-max procedure for action selection is assumed to allow for stochastic choice (due to unobserved factors). Applying a similar procedure here, letting A_i stand for agent *i*'s action set, with a_i as characteristic element, the probability P that action $a_i \in A_i$ is selected is represented by:

$$(8) \quad P_{it}(a_i) = \frac{e^{\theta_i V_{it}(a_i)}}{\sum_{h \in A_i} e^{\theta_i V_{it}(h)}}$$

(or V_{it}^{g} instead of V_{it} if faced with a generalized other), where θ_i is a nonnegative parameter, the inverse of which is labelled *temperature*. If $\theta_i \to 0$ each action is equally likely to be chosen, while if $\theta_i \to \infty$ the action maximizing utility is selected. Bault et al. (2017) provide a formal behavioral underpinning for this expression and also show how this decision model can be extended to allow for strategic intertemporal (future oriented) decision-making.

3 Support and some implications

Below scientific support for the *i*ATM model is presented coming from five different sources crossing multiple disciplinary boundaries: (1) indirect experimental evidence, (2) econometric evidence from a model-based statistical analysis of experimental data, (3) evidence from a model-based statistical analysis of brain scanning (fMRI) data, (4) further neurobiological evidence, and (5) translational and evolutionary evidence. In addition, some implications are discussed.

3.1 *Indirect experimental evidence*

The theoretical social ties model of van Dijk and van Winden (1992, 1997) assumes that an individual's care for another agent is determined by the accumulation of interaction experiences (impulses like in eq.(1)). Motivated by this model, several experimental studies investigated its general qualitative prediction – which also holds for the *i*ATM model – that the quality of interaction experiences will affect an individual's concern for the welfare of an interaction partner. In a first study, van Dijk et al. (2002) use a repeated public good game with fixed pairs for interaction and an adapted Ring-test of social value orientation (Liebrand 1984) to uncover changes in social preferences (care for the other). The social psychological concept of *social value orientation* (SVO, for short) stands for the weight an individual attaches to the utility of an anonymous, randomly selected other individual (a generalized other), and is typically tested in a social dilemma context (see, e.g., Van Lange 1999, Van Lange et al. 1997). It can be considered a practical measure of a generalized tie value (GTV) for such a context, as further discussed below. More specifically, and for later reference, the Ring-test involves a series of choices between two alternatives, with each alternative comprising a Self-Other payoff allocation. For example, alternative A may entail: +304 for Self and +397 for Other, while Alternative B may entail: +354 for Self and -354 for Other. Each payoff allocation represents a point on a circle, with zero payoffs as midpoint, and payoffs to Self (Other) on the horizontal (vertical) axis. Each payoff combination can be considered a vector, and the angle of the aggregate vector resulting from the summed preferred allocations as a measure of the individual's care for the other. The Ring-test measures an individual's SVO by having an anonymous, randomly selected other individual (doing the same task) stand for the

Other in the payoff allocations. In the same way, and for the same purpose, this task was applied in the experiment before the introduction and start of the interaction of the public good game. However, and importantly, after the interaction the task was applied again but now with the interaction partner serving as Other. As a consequence, the resulting angle of the aggregate payoff vector then gives a measure of the social tie with this specific interaction partner. The regression analysis of this study provides some qualitative support for the affective tie mechanism in that, in addition to a participant's SVO, interaction success measures appear to impact the tie with counterpart (more specifically, earnings in the last five rounds of the game before the second Ring-test). Furthermore, in case of an "individualist" (someone with a zero SVO score in the first Ring-test) a selfish-Nash equilibrium in the public good game would result in no social tie (a zero angle in the second test), suggesting that such behavior may have served as reference point. We will return to that later on. Moreover, their finding that participants adapted their contributions mostly in the same direction as the changes in counterpart's contributions is consistent with a continuous development of tie values (their analysis cannot exclude some form of strategic behavior, though). In addition, a debriefing questionnaire revealed that participants who said they would not like to continue with their present partner, in case of a repetition of the game, had developed a more negative differential between their tie value with this partner and their SVO than participants who answered they wanted to continue with their present partner. One result that the information-extraction based *i*ATM model may further help explain concerns their finding of a relatively strong impact of the SVO in the regression. According to this model, this may be due to less type-uncertainty in later rounds, with a relatively small impact of the impulses in these rounds as consequence (see eqs. (2) and (4)), while SVO picked up the impact of the early rounds (for the development of play in the different dyads, see Fig. C2 in Bault et al. 2017).

In a follow-up study, Sonnemans et al. (2006) investigate contributions to a public good in four-player groups instead of dyads. One of the main findings is that the tie value with a group member is again influenced by the SVO and the interaction success (contributions) with that specific other. Brandts et al. (2009) add to these findings by showing that emotions appear to mediate the tie value's dependence on interaction success (in this case, earnings in a two-player repeated social dilemma game). Instead of a repeated binary-choice Ring test, these two studies used a so-called Circle test, where a single point on a circle with payoff allocations to Self and Other had to be selected.

Further experimental evidence comes from two studies on the impact of gift giving. Malmendier and Schmidt (2012, 2017) find that an unconditional gift biases the subsequent decision of the receiver to the benefit of the gift-giver, even though there is no continuation of the interaction and the recipient understands the gift-giver's intention. Only by allowing that the gift can create a bond between the two agents – as in the social ties model – they can explain their results. In the model that they propose they introduce a weight attached to the utility of the interaction partner which depends on the partner's chosen strategy compared to an expected strategy. The latter strategy is like the reference action in the social ties model. The former strategy, however, concerns an intended action (profile), whereas the social ties model is about realized actions. Interestingly, a related experiment by Pan and Xiao (2016) shows that the gift receiver favors an actual gift over an intended gift, in line with the social ties model.

Consistent with many studies on the link between social value orientation and cooperation in social dilemma's (for a meta-analysis, see Balliet et al. 2009), van Dijk et al. find that the Ring-test measure of SVO correlates with the first few (but not all) decisions in their public good game experiment. According to the *i*ATM model, this SVO measure – as practical measure of GTV – should continuously adapt to interaction experiences, in contrast to the standard view of SVO reflecting a stable personal trait, that is, a stable preference weight (Bogaert et al. 2008, Van Lange 1999). Supportive of this predicted variability are findings of individual SVO (angle) changes after social interaction experiences in a social dilemma experiment (Brandts et al. 2009) and in a public good experiment (Ackermann and Murphy 2019), while van Dijk et al. (2002) find no impact of an individual decisionmaking task. Furthermore, the role played by contexts in the GTV construct finds support in the observed context dependency in the measurement of SVO (Greiff et al. 2016, Bogaert et al. 2008). Importantly, because the SVO literature typically focuses on a categorical distinction between SVO types (like prosocials versus individualistist) – comprising ranges of preference weights or angles as measured by the Ring-test, for example – high test-retest stability scores need not imply that the individual preference weights are stable.

3.2 *Econometric evidence*: *parameter estimates and predictive performance*

The above experimental findings provide some indirect support for the *i*ATM model, in particular regarding the influence of interaction experiences on affective tie formation, the GTV, and subsequent behavior. The findings presented next relate to more direct model-based econometric evidence.

The first econometric evidence comes from Bault et al. (2017). They adapted and extended the theoretical model of van Dijk and van Winden (1997) in mainly the following two ways. First, in their discrete (instead of continuous) time implementation of the model they incorporate an explicit linear specification of the tie mechanism, somewhat similar to eq. (2), and leave open the exact nature of the reference action in the impulse instead of taking the individual's own action as such (cf. eq. (1)). Although they suggest the possibility of doing so, their specification lacks the information-extraction based underpinning of the *i*ATM model. In contrast, they simply assume an exogenously fixed parameter (δ_{i1}) attached to the existing tie (labeled *tie-persistence* parameter) and another one (δ_{i2}) attached to the current impulse (labeled *tie-impulse* parameter). Second, they extend the model by allowing for forward-looking (strategic) behavior, using an additive two-period intertemporal utility function. Here, they assume that a decision-maker may expect an interaction partner to adapt her or his action choice in proportion to the differential between the decision-maker's own action and the other's expected action (with the proportion, β, labeled the *influence* parameter). For generating data fit for estimating the forward-looking model, they run a two-player public good game experiment where each round participant's expectation regarding other's contribution are elicited. In this game both the selfish-Nash contribution and the Pareto-optimal contribution are interior to the action space, making under- as well as overshooting of both possible. The following main findings are obtained by assuming soft-max action selection (similar to eq. (8)) and applying a maximum likelihood estimation procedure. *First*, all parameters of the myopic model ($\theta, \delta_{i1}, \delta_{i2}$) are significant, with both δ_{i1} and δ_{i2} being close to ½ after normalization of the impulse (as in eq. (1)). Notably, the finding that these parameters appear to add up to 1, based on *un*constrained estimation, actually supports the *i*ATM model, where $\delta_{i1} = 1 - \delta_{i2}$ (see eq. (2)); the estimate of ½, moreover, could be an average of a relatively high weight attached to impulses in early rounds and a lower one regarding later rounds (due to less type-uncertainty). *Second*, the forward-looking model, as specified, finds little support at the group level (insignificant β), which may be due to heterogeneity. Mixture model estimation – using individual estimation for classifying participants first as myopic or forward-looking – suggests that about one-third of the participants are looking forward (with $\beta = 0.3$). All in all, this second finding is supportive of the relevance of the (myopic) affective social ties submodel. *Third*, as reference action, the selfish-Nash contribution level shows the best performance (based on likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) as measures of fit), compared to a number of other potential reference points, including the other's expected contribution (cf. the

Malmendier and Schmidt model, referred to above). *Fourth*, compared to a prior tie value of 0 – which may seem natural as the interaction in the experiment is with an anonymous individual – taking the SVO (measured via a Ring-test, without feedback, as in van Dijk et al. 2002) improves the fit of the data. As operationalization of a GTV, this provides further support for the related third module of the *i*ATM model.

In a follow-up experimental study Loerakker et al. (2016) construct a Fragile Public Good game to allow more room for harmful behavior and negative tie formation. In this repeated game, with fixed dyads, participants start each round with some tokens (exchangeable for money) in a private account and in a common account. Contributing tokens to the common account or taking tokens from it is fully symmetric in terms of own costs and benefits. The only difference between contributing and taking is that the former helps and the latter hurts (with the same amount) the other participant one is matched with. Again, both the selfish-Nash and the Pareto-optimal action are interior to the action space, with the former now being equal to the status quo (at the start of each round). Also the forward-looking part of the theoretical model, the use of the Ring-test for SVO, and the estimation procedure are as in Bault et al. (2017). Here, the most relevant findings are that there is again support for the myopic model (with again, at the individual level, about one-third of the participants looking forward) and that both δ_{i1} and δ_{i2} are close to ½ after normalization of the impulse, in line with the *i*ATM model. Unexpectedly, they also find some evidence of a stronger response to a positive compared to a negative impulse that cannot be explained by their background social ties model. The *i*ATM model suggests a potential explanation, though. Due to the plausibly more threatening equally ample space for negative actions in this game, greater uncertainty and a more positively skewed type-distribution may have been induced, making positive impulses more surprising (besides more unlikely, also more unexpected), with a larger weight attached to positive impulses as consequence (see, e.g., Faraji et al. 2018, Liakoni et al. 2021).

Both Bault et al. (2017) and Loerakker et al. (2016) further look at the predictive performance (fit) of the estimated model both within and across different datasets and in comparison with other extant models. Because of data availability, and in view of its relatively good performance, they use the myopic version of the model and a prior tie value of 0. Bault et al. use van Dijk et al. (2002) as alternative dataset, and further compare the fit of the endogenous affective ties model with a standard fixed social preference (α) model, and an inequality aversion model (Fehr and Schmidt 1999). Loerakker et al. use a subsequent repeated game in their experiment (with another anonymous participant) for out-of-sample prediction, comparing the fit of the ties model with a fixed social preference model and the inequality aversion model, and, in addition, a reinforcement learning model (Roth and Erev 1995). The results provide further support for the affective ties model. Parameter estimates obtained from the alternative dataset of van Dijk et al. (2002) are very similar, and, compared to all the other relevant models, the fit of the estimated affective ties model comes out best (with both group-level as well as individual-level estimates).

3.3 *Model-based fMRI evidence*

An interesting new source of potential evidence is provided by neural data. Encouraged by the good behavioral fit of their estimated social ties model, Bault et al. (2015) carried out a model-based brain scanning (fMRI) study, linking their parameter estimates to brain activity, to see whether a neural substrate exists for the postulated affective tie mechanism (see previous item). Because the extended forward-looking two-period model did not perform better, they use the simpler myopic model in their analysis. Their main findings are: (1) during the contribution decision phase, activity in the posterior superior temporal sulcus (pSTS), temporo-parietal junction (TPJ), posterior cingulate cortex (PCC) and several areas in the frontal lobe (among which the medial prefrontal cortex (mPFC)) show a *negative* parametric modulation by the *tie value* (α_{ijt}); (2) during the feedback phase, activity in the pSTS, TPJ, and the Insula, show a *positive* parametric modulation by the *impulse* (the other's contribution minus the reference contribution); (3) during the decision phase, activity in the mPFC, PCC, and Insula show a *positive* parametric modulation by the *contribution magnitude*; (4) a *connectivity* analysis reveals a significant increase in correlations between the activity of the pSTS and activity of the mPFC and PCC during the decision phase (PPI analysis, masking results from the contribution parametric effect); moreover, (trialwise) parameter estimates of the tie-parametric effect on pSTS activity at the start of the decision phase significantly correlate with parameter estimates of the contribution parametric effect on mPFC activity during the decision validation phase (beta-seed correlation analysis); (5) during the decision phase, δ_{i1} (the weight attached to the existing tie value in tie formation) correlates with activity in the right TPJ and right pSTS, and δ_{i2} (the weight attached to the impulse in tie formation) correlates with activity in the left TPJ. Together, these findings are consistent with the existence of a neural substrate for the various components of the model, regarding tie formation and decision-making.

This interpretation is corroborated by substantial neural evidence regarding the functioning of, in particular, the TPJ/pSTS and mPFC. The TPJ and its neighboring pSTS are consistently activated during social interaction. Studies show that this brain area is implicated in inferring beliefs and intentions of others, tracking another agent's strategies and influence updating, the reliability or behavioral relevance of someone, liking ratings, other's cooperativeness, as well as attentional shifts (see references in Bault et al. 2015). Other studies relate this area's activation to trust (Engelmann et al. 2019), altruism (Morishima et al. 2012, Tankersley et al. 2007), closeness (described as inclusion of other in the self; Cheng et al. 2010), social discounting (Strombach et al. 2015), and overcoming self-centeredness (Soutschek et al. 2016). The very different tasks involved make it difficult to determine the underlying framework of computations that are performed. Geng and Vossel (2013) propose "contextual updating" as framework: "the purpose of this area is to update internal models of the environment (including other people) for the purpose of constructing appropriate expectations and responses" (ibid., p2617). The neural finding of Bault et al. regarding the social ties model, suggesting a computational role of this brain area in encoding the significance (or type in the *i*ATM model) of other people as a resource for goal fulfilment, is consistent with this view. That is, granted that the internal model updating includes the appropriate valuation of those other people (formalized as a social preference weight). The latter distinguishes the social ties (and *i*ATM) model from a standard reinforcement learning model, as it concerns the nature of the valuation function itself rather than the value of a choice option for a given valuation function.

Regarding the mPFC, substantial evidence exists suggesting that this area is implicated in tracking the predicted (nonsocial and/or social) reward associated with a particular choice, while previous studies have also reported the functional connectivity between the pSTS/TPJ and the mPFC during decision-making (see references in Bault et al. 2015, Hill et al. 2017). What makes the above neural findings particularly interesting is the reported negative correlation between the estimated tie value and activity in the pSTS as well as the functionally connected mPFC activity during decision-making, and the support for the claim that the signal shared between these two areas contains information regarding the tie value (based on beta-seed correlation). The former makes sense as the (strategic) balancing of own and other's benefits and costs and controlling the appropriate (non-selfish) response becomes less effortful the closer to the self the other agent is regarded (see also Strombach et al.

2015), while the latter finding plausibly suggests that the mPFC activity per choice (here, contribution) unit becomes smaller with greater closeness.

Finally, regarding the parameters δ_{i1} and δ_{i2}, note that in the model of Bault et al. these parameters are considered to be stable personality traits, with δ_{i1} (labeled tie-persistence parameter) assumed to be revealing the speed with which the tie deteriorates over time if the interaction is not maintained, and δ_{i2} the impact of counterpart's behavior on the new tie value. In contrast, the *i*ATM model suggests that these parameters are not only complementary (adding up to 1) but also endogenous, with as an important implication that δ_{i2} becomes smaller – and thus also the impact of the other's behavior (the impulse) – the better counterpart's type is known (for supportive evidence, see Section 4). A common experience like nothing has changed upon reunion with an old friend may count as anecdotal evidence that a tie may not deteriorate, but leaves a state that remains eligibile for updating (an eligibility trace; see Niv in Glimcher and Fehr 2014, pp305-306). What changes is that memory retrieval may become more effortful (and, thus, more costly) the longer a tie has not been maintained.

3.4 *Further neurobiological evidence and modeling*

The behavioral and neural support for the affective tie mechanism summarized above raises the question how these findings relate to the recent neurobiological literature on maternal care (attachment) and bonding of mammals (Insel and Young 2001, Numan 2015, 2016, Numan and Young 2016, Feldman 2016, 2017). The aim of this item is to show that the observed role of the pSTS/TPJ in affective tie formation can be naturally fitted into an emerging neural network model of human bonding. This network model builds on the maternal care system, which is suggested to provide the neural foundation of human bonding more generally (as in friendships; e.g., Numan 2015, p271). The evidence-based network model proposed by Numan (2016, 2015) will serve as workhorse (see also Numan and Young 2016). As it turns out, the *i*ATM model may be even more directly related to this pSTS-enriched neural network model.

Numan's hypothetical neural network model of maternal care

Starting point in Numan's hypothetical network model of maternal care – largely based on evidence coming from nonhuman mammals – is the basic and automatic subcortical reward system. Infant stimuli are received as input by two key brain areas, the medial preoptic area (MPOA, part of the hypothalamus) and the amygdala. These stimuli may activate positively valent (prosocial) or negatively valent (antisocial) neuronal circuits, dependent on whether they are perceived as beneficial or harmful/aversive. Interestingly, nulliparous females typically avoid infants, because infant stimuli activate their antisocial circuit. The jumpstart for maternal care in this network model is provided by pregnancy hormones prolactin and estradiol and the peptide oxytocin acting on the MPOA. MPOA output inhibits the activation of the antisocial circuit (in the amygdala and other parts of the hypothalamus), and activates the mesolimbic reward system in a way that the stimuli become motivational (attractive in this case). The latter happens by activating dopamine (DA) neurons in the midbrain ventral tegmental area (VTA), which stimulates dopamine release into the nucleus accumbens (ventral part of the striatum). This causes its inhibition of the ventral pallidum (with the striatum part of the basal ganglia) to be released, allowing the ventral pallidum to become responsive to the prosocial neuronal output of the amygdala, with approach behavior (attraction) towards the infant as consequence. Importantly, the MPOA also stimulates the release of oxytocin (OT) by the paraventricular nucleus (PVN, another part of the hypothalamus) in the various brain sites discussed. The interaction between OT and DA along the circuitry is considered to be critical to the effects just mentioned (I will return to this below). In case of negative social stimuli, evidence suggests that, apart from negatively valent neurons in the MPOA and amygdala, negative neural pathways implicating additional parts of the hypothalamus and now the periaqueductal grey (PAG, a midbrain pre-motor area) are involved in reflexive antisocial fight (approach) or flight (avoidance) responses. Alternatively, more proactive and goal-directed antisocial responses like voluntary withdrawal (avoidance) or spite (approach) appear to be possible through projections of the PAG towards the VTA and the subsequent activation of negatively valent pathways in the nucleus accumbens and ventral pallidum.

Continuing with (positive) maternal care, after the initial "recognition stage", a stage of "persistent attraction" is explained by the strengthening of synapses (neural plasticity) between the relevant neurons in the amygdala and the ventral pallidum. This enables continuation of maternal behavior after the hormone induced onset has faded. To account for feeling states in humans (maternal love and empathy), Numan extends this hypothetical neural model with additional links between the amygdala and two cortical limbic brain areas implicated in emotions: the orbital frontal cortex (OFC) and insular cortex (IC). In turn, these areas are linked with the MPOA through the medial prefrontal cortex (mPFC), and thereby linked with the above discussed circuitry for maternal care behaviors.

Because males are not similarly exposed to pregnancy hormones, paternal care cannot be explained along the same lines. In this case, Numan suggests that interaction experience with a pregnant partner and subsequently with the infant may engage the same neural network, activated by experience-induced endogenous oxytocin. Interestingly, examining brain responses to infant cues, Abraham et al. (2014) find higher STS activation in care-giving fathers compared to substantially stronger amygdala activation in mothers, with STS activity being associated with OT, and the degree of amygdala – STS connectivity in fathers being related to the time spent in direct child care.

Mapping the affective tie mechanism onto the neural network model

These findings provide a stepping stone for including the neural findings regarding the affective tie mechanism discussed under the previous item, in particular, because of the observed role of interaction experience and the connectivity between the STS, the mPFC, and the amygdala (see also Decety and Svetlova 2012, Bickart et al. 2014, Pitcher et al. 2017), with potential modulation of the pSTS by OT (Bethlehem et al. 2013, Gordon et al. 2013, Abraham et al. 2014). As empathic concern seems not automatic but dependent on the valuation of the other (Singer 2006, Batson et al. 2007, Hein et al. 2010, Decety and Svetlova 2012; see also Fahrenfort et al. 2012), the following slight adaptation of the Numan network model is proposed. In addition to having the empathy related brain areas provide a link between the amygdala and the mPFC, another one would be provided by the pSTS as integrator of information (memories) concerning counterpart and context. Given an affective tie, empathy (embodied simulation, Feldman 2017) may be expected to play an important role in appropriate caring, as it requires an assessment of the effect of one's behavior on the welfare (utility) of the one cared for. In the affective ties model, this assessment gets formalized through the specification of other's utility in the extended utility function (eq. (5) in Section 2). The aforementioned neuronal circuits would facilitate such computations by the mPFC. Interestingly, not only the mPFC but also the Insula (IC, implicated in empathy) showed a positive parametric modulation by the contribution magnitude during the decision phase of the experiment of Bault et al. (2015) discussed above.

The affective ties model helps explain why not only the direct interaction experience of a father with his infant but also the interaction with a beloved (pregnant) partner may induce paternal care. The reason is that the care would be appreciated by the partner, which increases the father's extended utility. Furthermore, note the similarity between the notion of "persistent attraction" and a positive existing tie value. Consequently, the affective ties model would predict that even if an infant stimulus would be perceived as negative it need not turn a mother's positive care into negative care. More generally, regarding approach versus avoidance, the affective ties model helps distinguish between different cases. Because the overall model concerns goal-directed behavior, more reflexive responses are neglected for the moment. In case of a positive tie value, prosocial and approach behavior is stimulated. In case of a negative tie value, antisocial behavior is stimulated, but this need not always entail avoidance. Approach in that case would occur if a costly spiteful action, aimed at hurting counterpart, would predict a net-benefit, reflected by a positive first-order derivative of the agent's extended utility function. Extremely high (absolute) tie values may induce mere reflexive behavior, like the textbook fight (approach) or flight (avoidance) response. In that event the more direct amygdala-hypothalamus-PAG path, discussed above, would become relevant.

A speculative neural underpinning of the *i*ATM model

The *i*ATM model may be even more directly related to the pSTS-enriched neural network model of Numan. Though the following neural underpinning is admittedly speculative and simplistic, it may open up a potentially important perspective of a multi-level bonding model bridging various disciplines (see also the next item on translational and evolutionary evidence). Starting point are the following four observations in the recent literature. *Firstly*, OT is found to be not only involved in facilitating prosocial behavior, as often suggested, but also in facilitating antisocial behavior (see, e.g., De Dreu 2012, Olff et al. 2013, Guzmán et al. 2013, Kemp and Guastella 2011, Kelly and Vitousek 2017,). *Secondly*, according to Bartz et al. (2010) OT seems to orientate attention to social stimuli and facilitates the encoding of social memories (of interaction experiences) along with the hedonic value of the stimulus (impulse). *Thirdly*, Churchland and Winkielman (2012) suggest that the anxiolytic effects of OT can explain the majority of findings. And, *fourthly*, OT appears to be involved in social as well as nonsocial physiological and behavioral responses in adaptation to changing environments (Feldman et al. 2016, Quintana and Guastella 2020). Together these observations lead to the conjecture that the signaling molecule OT is involved in the following: orientating attention to (social or nonsocial) stimuli related to environmental uncertainty, reducing that uncertainty through information extraction, and facilitating the encoding of stimulus and context related memories (interaction experiences) including the hedonic value of the stimulus (impulse). This may help to further clarify the OT-DA interaction in Numan's neural network model. Whereas DA is related to reward prediction and reward prediction error, irrespective of the stimulus type (Schultz et al. 1997, Montague et al. 1996), the role of OT would seem to be related to stimulus (source) type prediction and type prediction error. OT-DA interaction would then facilitate the factoring in of the type assessment into the computation of reward in the striatum, where neuronal activity appears to reflect action values for self and other, instrumental in the preparation of decisions (Báez-Mendoza and Schultz 2013). To illustrate, how the *i*ATM model might fit into this picture, assume again a (two-person) public good game context, and let it start for protagonist *i* with an impulse I_{ijt} as stimulus and a prior tie value α_{ijt} (initially equal to GTV_{it}), represented by the activity of neurons in the amygdala and MPOA. The impulse I_{ijt} triggers a prediction error, made up by the difference between the impulse and the prior $\left(I_{ijt} - \alpha_{ijt}\right)$, which is encoded by both brain areas, facilitated by the related MPOA-instigated release of OT from the PVN. Dependent on the sign of the prediction error, positively or negatively valent neurons in the amygdala are activated. The prediction error is further communicated to the pSTS (and neighboring TPJ), where assumedly the activity of a population of neurons reflecting the type distribution (and the uncertainty related learning rate δ_{ijt}, see Subsection 3.3) is adjusted via neural plasticity, generating a new tie value α_{ijt+1} (eq. (2) in Section 2). This updated tie value is subsequently fed forward to the mPFC for decision-making, in case of a repeated interaction within the same context. In preparation of the decision, the mPFC would then inform the amygdala and MPOA of the relevant new tie value as type prediction and new prior (a form of predictive coding (Brown and Brüne 2012)). Facilitated by related OT release from the PVN, this internal stimulus may set in motion the striatum-assisted reward computation of the network model, with further empathic input from IC, leading to prosocial or antisocial behavior (recall that a tie value generates a social preference weight in the model). In case of a new counterpart or a new context the mPFC would assumedly retrieve an appropriate GTV_{it} as prior from the pSTS.

Other brain areas and neuromodulators

Of course, a more complete model of human decision-making would require attention for additional brain systems. This would include brain areas implicated in episodic memories, for example, important for recognition and GTV computation (in particular, the medial temporal lobe), and systems involved in higher-order cognition for strategic deliberations (requiring input from the more dorsal and lateral prefrontal cortex); see, e.g., LaBar and Cabeza (2006), Lengyel and Dayan (2007), Shohamy and Wagner (2008), Yonelinas and Ritchey (2015). Furthermore, although oxytocin has attracted most attention, other important neuromodulators, like vasopressin, serotonin, and opioids should be accounted for in a more complete model (Dunbar 2010, Numan 2015, Numan and Young 2016, Declerck et al. 2013, Feldman 2017).

3.5 *Translational and evolutionary evidence*

Evidence on prosocial behavior and enduring relationships among very different animal species, and even plants and bacteria, suggests the potential relevance and applicability of the *i*ATM model to a much wider range of organisms, as may be expected given the ubiquitous challenge faced by organisms of adapting to behavioral uncertainty. Before proceeding it may be useful to repeat the two key characteristics of the affective tie mechanism: (1) the cumulative assessment and encoding of the experienced beneficial or harmful behavior of another agent, generating an estimate of it's friend or foe *type*; (2) in turn, this type assessment induces *care* for that other agent, where care stands for the positive or negative valuation of its welfare. Neither the encoding nor the caring needs to be conscious. Moreover, a tie may be specific or generalized, while context and attention play a role (cf. Cronin 2012).

Enduring relationships among animals, plants, and bacteria

Examples abound of animals showing prosocial behavior – improving another individual's welfare – and enduring relationships (partnerships, bonds, sometimes called friendships; see Massen et al. 2010, Seyfarth and Cheney 2012). Among them are primates like chimpanzees and baboons, horses, dolphins, elephants, hyenas (for references, see Seyfarth and Cheney 2012), cows (e.g., de Freslon et al. 2020), rodents such as voles (Young and Wang 2004) and rats (Ben-Ami Bartal et al. 2014), birds like parrots (Brucks and von Bayern 2020), and fish (Soares et al. 2012). Evidence of prosocial behavior and enduring relationships extends to plants and bacteria, as indicated by studies on mutualisms between plants and mycorrhizal fungi (Kummel and Salant 2006, Kiers et al. 2011, Fellbaum et al. 2012) and between legumes and rhizobia (Simms and Taylor 2002, West et al. 2002). Mutualisms are reciprocally beneficial relationships or interactions, where an organism performs a behavior (usually with some short-term cost) that provides a benefit for an individual of a different species (West et al. 2002).

Importantly, these relationships are not only based on providing useful resources but may also involve negative sanctions (e.g., withdrawal) in case of harmful behavior (West et al. 2002, Kiers et al. 2011). These mutualisms show that interaction-based prosocial behavior need not even involve (related) conspecifics. The same holds for animals. To give an example, in a recent study (Ben-Ami Bartal et al. 2014) rats helped trapped strangers (just as cage mates) by releasing them from a restrainer, whether they were of their own strain or not. In case of a different strain they only did so, however, if they had been previously housed (and, thus, had experience) with the trapped rat. Furthermore, pair-housing with a rat of a different strain prompted rats to help strangers of that strain. Moreover, rats fostered from birth with another strain, and not their own strain, helped strangers of the fostering strain but not rats of their own strain. This clearly shows the importance of social experience (familiarity) for prosocial behavior and provides evidence against an innate bias. Ben-Ami Bartal et al. (2014, pp 9-10) conclude that "through social interactions rats form affective bonds that elicit empathy and motivate helping. This motivation to help is extended to strangers of familiar strain."

The above evidence seems consistent with the conceptualization of the affective tie mechanism as a proximate mechanism, suggesting the potential usefulness of the *i*ATM model, although the precise way of type-encoding (information extraction and integration) and caring (counterpart valuation) may be different, and more or less sophisticated (think of empathy, for instance) for different organisms. Whereas the literature has typically focused on ultimate mechanisms of altruism and cooperation (see below), more recently an interest has grown in underlying proximate mechanisms of (costly) prosocial behavior and partnerships. In their literature review, focusing on (non-kin) primates, Schino and Aureli (2009) argue in favor of an "emotional bookkeeping" system that appears to be quite similar to the affective tie mechanism (except for lacking a formal model). Their argumentation goes as follows. Although altruistic or (costly) prosocial behavior may be favored by selection because of subsequent benefits, it does not follow that such behavior is (proximately) motivated by these future benefits, that is, by the expectation of return favors (ibid., p53). In view of the limited cognitive skills of many animals the assumption that they plan social interactions to obtain future benefits may well be unwarranted. Proximate mechanisms assuming that animals are motivated by previous, rather than future, benefits may be favored by natural selection because past behavior is often predictive of future behavior (ibid., p54). Moreover, through the flexibility of partner choice, mistakes need not be very costly. What is needed is a partner-specific "memory" of the benefits received; an episodic memory is not needed, as the formation of an emotional bond can suffice (ibid., p55). In short, the idea is that: "the exchange of services triggers partner-specific emotional variations, and that animals make their behavioral decisions on the basis of emotional states associated with each potential partner. The development of differential social bonds with individual group mates, thus, corresponds to an emotionally based bookkeeping system of received services in which emotions provide the basis for "rules of thumb" that guide social choices." (ibid., p59) They note that emotional mediation makes long-term reciprocity possible (cf. Brosnan and de Waal 2002) and that it allows for the conversion of the value of different behavioral episodes (services like grooming or food sharing) into a common currency. At least for primates this "emotional bookkeeping" approach seems relevant (see also van Hooff 2001, Aureli and Schaffner 2002, Schino and Aureli 2010, Evers et al. 2015, 2016), and shows clear parallels with the affective tie formation part of the *i*ATM model.

Of course, in the context of plants and bacteria one should think of emotions and affect in an appropriate way. Appraisal theory of emotion (see, e.g., Lazarus 1991, Scherer et al. 2001, Frijda 2007) offers some leeway for translational continuity (cf. de Waal 2008) by viewing emotions as being determined by the evaluation (appraisal or estimate) of an event or behavioral episode, which can be more or less refined, and does not need to involve consciousness (cf. Aureli and Schaffner 2002). This approach would seem to be able to accommodate the behavior of even relatively simple organisms. From this perspective, the type assessment part of the *i*ATM model may be seen as the formalization of an emotional appraisal process concerning the helpful or harmful behavior of a counterpart, facilitating a parsimonious behavioral model. As "affective" in the concept of an affective tie mechanism relates to the subsequent taking into account (valuing) of that counterpart's welfare (the "caring" part), in principle, also this key characteristic of the model would seem to be applicable to the behavior of simpler organisms.

Considerable continuity across species also appears to hold from a physiological perspective. Oxytocin-like peptide signaling systems appear to be more than 600 million years old (Grimmelikhuijzen and Hauser 2012, Gruber 2014, Feldman et al. 2016, Quintana and Guastella 2020). These peptides – and the same holds for structurally similar vasopressin – presumably evolved from ancestral vasotocin and are present in vertebrates, including mammals, birds, reptiles, amphibians and fish. They have been identified also in invertebrate species, such as nematodes and arthropods, and it seems that these signaling systems have conserved functions in physiology, including water homeostasis, reproductive behavior (such as mate recognition), learning and memory (Gruber 2014). Therefore, it is expected that they are related to the formation and maintenance of affiliative social relationships in many animal species (for some evidence, see Massen et al. 2010). Quintana and Guastella (2020) more generally argue that oxytocin is best described as an "allostatic" hormone, facilitating the adjustment of sensing and response set-points, assisting learning and prediction to better adapt to changing environments, which is crucial for survival. Note that, consistent with this view, the dynamic friend or foe type estimate in the *i*ATM model (the tie value) similarly functions as a dynamic response set-point, involved in an environmental learning and prediction process. Finally, recent findings in the new field of "plant neurobiology" suggest that similar signaling hormones in plants may play a role in their behavioral plasticity and sociality with other plants or other organisms (Brenner et al. 2006, Baluška et al. 2006).

All in all, it seems that a formal theoretical model like the environmental uncertainty based *i*ATM model may be more widely applicable to animal, and perhaps even plant behavior. The advantages of having such a formal model are, among others: greater precision (e.g., regarding the temporal sequence of behavioral events), organization of results (potentially across different species), facilitation of predictions (think of the correlational evidence problem) and of new hypotheses. Furthermore, it may offer (alternative) explanations. To give one example, Tennie et al. (2016) find no evidence of helping by chimpanzees in their experiment and suggest that findings of prosocial behavior may be a by-product of task design. Although their latter point is well taken, the *i*ATM model suggests an alternative explanation for no helping behavior. Because they made effort to minimize the effects of personal relationships (and recipients could not respond), according to the *i*ATM model, each chimpanzee's (generalized) tie value may well have been zero, approximately. In that case, the model predicts no helping, as observed. With (repeated) interaction, prosocial (or antisocial) behavior might have shown up, though.

*i*ATM from an ultimate mechanism perspective

The above shows that the *i*ATM model, as a proximate mechanism for bonding and prosocial (or antisocial) behavior, finds substantial support from the behavioral and life sciences. Although it is beyond the scope of this paper to thoroughly discuss the *i*ATM from an ultimate mechanism perspective, a few remarks are in order. First of all, it can induce tit-for-tat (TFT) resembling reciprocity in a prisoner's dilemma setting – a type of behavior that can be evolutionary stable in certain environments (Axelrod and Hamilton 1981, Nowak 2006). And, it also helps explain experimentally observed behavioral adaptation to benefit-to-cost ratio changes in repeated prisoner's dilemma games (see Loerakker et al. 2016). Apart from reciprocity, *i*ATM seems consistent with various other rules for the evolution of cooperation distinguished in the literature (see, e.g., Nowak 2006, Bowles and Gintis 2011, Kramer and Meunier 2016). For example, Hamilton's rule, regarding inclusive fitness or kin selection as ultimate reason for an altruistic act, requires that the degree of relatedness (r) should exceed the cost-to-benefit ratio (*c/b*): $r > c/b$. Instead, the *i*ATM model requires: $\alpha > c/b$. In general, there is no reason to expect that the degree of relatedness will equal the tie value (i.e., $r = \alpha$). However, in mammals, due to the bonding between parental caretakers and infants, the potential indirect ties with any other relatives (such as other siblings) through the ties with parents, and the affective-tie related proximity seeking facilitating further tie formation, this equality may be approximated, at least in a directional sense. Note that tie-related cooperation can be altruistic (costly) from an selfish-utility point of view (but not from an extended utility viewpoint), and is not innate, nor necessarily restricted to kin. Furthermore, through its potential of internalizing external effects of behavior by caring, affective tie formation facilitates the production of public goods (such as defense against threats from nature or other social groups) which plays a prominent role in group selection theories. In fact, the affective tie mechanism binds together the formation of groups and their internal cooperation that the evolution of sociality appears to require, but that are typically studied as seperate themes (van Veelen et al. 2010). The context dependency of the GTV, moreover, fits the view that there is no best rule independent of the environment (Axelrod 1980).

Finally, some remarks are in order on the start of tie formation if no ties already exist. Whereas with TFT it is typically assumed that it starts with cooperation, the *i*ATM model would seem to predict no cooperative behavior without any ties. However, several factors might generate cooperative actions and affective tie formation in the initial absence of affective ties. First of all, recall that the model assumes stochastic decision-making, Therefore, cooperative actions may happen, which may trigger tie formation leading to mutual cooperation, particularly if the choice space is lumpy (as in a standard binary choice prisoner's dilemma game). Second, cooperative actions may be a by-product of optimal self-oriented behavior, for example, when a positive contribution to a public good is optimal from an agent's own utility perspective (as in case of the leaky bacterial functions benefiting other bacteria in Morris et al. (2012), or the local public goods model of van Dijk and van Winden (1997)). Note, however, that a subsequent cooperative response would not be a by-product if produced by a resulting affective tie (in contrast with the by-product mutualism considered by Morris et al.). Third, according to the *i*ATM model, a positively skewed type distribution as prior in case of a threatening environment may turn a neutral action into a positive impulse generating a positive tie and relatedly cooperative behavior. Fourth, pregnancy hormones and opioids facilitate tie formation *ab ovo* in mammals. Note, finally, that internalized norms for cooperation are unlikely to play a role in this context, as a positive valuation of the norm sender (educator), implying a positive tie, seems necessary for successful internalization, generating an intrinsic motivation (cf. Pedersen 2004, see also item (6) of the next section).

Organismality and the primordial caregiving system

To conclude this item, two further topics are shortly addressed: organismality and maternal behavior as the primordial caregiving system. First, consider the concept of *organismality*, proposed in biology as a generalization of organism. According to Queller and Strassmann (2009, p3144): "The most salient feature of organisms is adaptation, the seeming goal-directedness that makes organisms different from merely physical entities. (…) We suggest that the essence of organismality lies in this shared purpose; the parts work together for the integrated whole, with high cooperation and very low conflict. Specifically, the organism is the largest unit of near-unanimous design; the qualifying 'near' is required because some conflicts (…) probably remain in all organisms. (…) We believe that organisms should be defined by what they actually do (…)" (see also the supportive review of West and Kiers 2009). The authors use a two-dimensional space, spanned by the extent of cooperation and the extent of conflict, explored at different levels of sociality: groups of cells (e.g., a bacterial biofilm or a mouse), groups of multi-cellular individuals (e.g., a honeybee colony or a human city), and multi-species groups (e.g., fig-wasp or legume-rhizobium associations). The extended utility formalization of the *i*ATM model (see eq. (5) in Section 2) seems to offer a fruitful formalization. Using the above definition of organismality and focusing, for simplicity, on a bilateral tie between individuals *i* and *j* (assuming type-risk neutrality and no informational conflict regarding utilities), their organismality may be positively related to the average of their tie values: $(\alpha_{ij} + \alpha_{ji})/2$, as measure of their cooperation level, and negatively related to the absolute difference in these values: $|\alpha_{ij} - \alpha_{ji}|/2$, as measure of their conflict level. Thus, if $\alpha_{ij} = \alpha_{ji} = 0$ their organismality would be 0, while its measure would equal 1 if $\alpha_{ij} = \alpha_{ji} = 1$. In the latter case, both individuals would persue the same goal (maximizing their joint utility), with no conflict. Clearly, if their perception of each other's utility differs, a distinction would arise between the organismality as seen from within (i.e., by an individual) and from the outside (by an informed observer of the two) (cf. Aureli and Schino 2019). Moreover, the contextual nature of extended utility in the *i*ATM model is consistent with the concept of "contextual organismality" proposed by Diaz-Muñoz et al. (2016), who refer, for example, to bacteria increasing their cooperative interactions in case of nutrient starvation to create a fruiting body, a new form of organismality.

The second and last topic to be addressed concerns the idea that maternal behavior is the primordial caregiving system and that, consequently, the neural systems underlying maternal behavior may have served as a foundation for other types of prosocial bonding (see, e.g., Numan 2015, p271, including references). If the above argumentation regarding the evolutionary origin of the affective tie caring mechanism is correct, it raises the question whether mother-infant bonding in fact piggybacked on this more fundamental affective tie mechanism, assisted by pregnancy hormones and opioids to provide a jumpstart for attachment with the fetal allograft (see Nelson and Panksepp 1998, Douglas and Russell 2001).

4 Applications

In summary, the informational affective tie mechanism (*i*ATM) concerns a fundamental agent-type information extraction route to caring, which is automatic and impulsive (non-deliberative) and is triggered by another agent's behavior. In humans it is distinguishable in terms of brain activity from higher-order mental processes that may lead to similar behavior, such as internalized-norm satisfaction or the calculus of reciprocity. Importantly, care for another agent need not always be positive but may involve spite or hatred, a negative social preference inducing antisocial behavior. Furthermore, the impact of the *i*ATM (a System I phenomenon) on decision-making may be influenced by higher-order (System II) cognitive processes, such as self-control (the regulation of emotional urges). Applications of the *i*ATM are discussed next, subsumed under 10 items.

(1) *Relative influence of the iATM on decision-making*: *time pressure and cognitive load*

Because of the primacy of affect, it may be expected that the relative influence of the *i*ATM on decision-making increases if there is less time to decide, that is, under *time pressure*. The model provides an information-based explanation for the finding in a meta-study of (mostly) one-shot public good game experiments that time pressure increases cooperation (Rand et al. 2014). The Social Heuristic Hypothesis put forward in that study suggests that this is due to prior positive experiences with cooperation that become generalized to the lab situation as an intuitive heuristic (see Rand and Epstein 2014, for field evidence). Different experiences are likely to moderate this effect of intuition relative to reflection (working against cooperation in a one-shot game), which helps explain the variance observed in the data. According to the *i*ATM model the time-pressure effect of more cooperation in the one-shot public good game may be explained by a positive generalized tie value (GTV) among participants, generated by prior behavioral experiences in similar contexts, weighted by the attention they receive in the decision-making process. An increase in cooperation need not always be observed, however, because it would depend on a sufficient number of participants having a sufficiently strong GTV > 0. Moreover, the prediction of the paper that “intuition should never decrease average cooperation relative to reflection” (ibid., p2) need not hold, because a GTV < 0, motivating harmful behavior, could establish just that if action space allows.

The *i*ATM is likely to have a stronger impact also if fewer mental resource are available due to *cognitive load*. The greater role of participants’ GTVs in that case, jointly with a variance in GTV values, helps explain why applying cognitive load to dictators in a dictator game experiment increases giving in some but not all experiments (Cornelissen et al. 2011, Hauge et al. 2016). Interestingly, Cornelissen et al. find that this effect is mediated by perceived interpersonal closeness, with the more generous prosocials perceiving random other people to be closer to them than proselfs. They use a social value orientation task to distinguish between prosocials and proselfs, which gives a practical measure of an individual’s GTV (see Section 3). As affective ties are related to interpersonal closeness, their finding is consistent with the greater role of the GTV suggested above.

(2) *Importance of interaction, social preference drift, and tipping points*

As impulses, driving affective tie formation in the *i*ATM model, are related to interaction experiences, *interaction is key* following this model. Consequently, this mechanism cannot work if interaction is thwarted through the building of walls between people, also if these walls concern ideological in-group versus out-group barriers. Regarding the latter, substantial evidence exists of the importance of bridging affective ties (friendships) in overcoming negative attitudes between such groups (Pettigrew 1998, Pettigrew and Tropp 2006). The reason can be twofold. One may behave towards a friend of a friend more positively because it would be appreciated by one’s friend, who’s welfare is part of one’s extended utility, but also if one observes that the behavior of the friend’s friend increases the welfare of one’s friend. The former is related to the presence of an indirect tie via one’s friend, and the latter to the development of a direct tie with the friend’s friend (see further item (9) below on identity). Interaction should not be narrowly interpreted, though, as in some cases the experience of no action (thus, no harm) may generate some positive bonding, and impulses may be a by-product of self-serving behavior (see the previous section). This leads to the next issue. Whereas standard social preferences models focus on given preferences and/or rational beliefs about the preferences or intended strategies of counterparts, the *i*ATM model focuses on (automatic) emotional appraisals of experienced behavior (recall the evidence regarding gift-exchange in the previous section). Although we do not want to argue against the potential importance of higher-order cognitive reasoning skills, many findings show that the evolutionary more primitive affective tie mechanism can still have substantial bite. Especially, because there is often little time, training or experience regarding a social decision problem; and, even if that does not hold, emotional arousal may thwart rational decision-making (for references and further discussion, see: Bault et al. 2017, van Winden 2015).

Another consequence is that the affective tie formation through social interaction will entail a *social preference drift*. Furthermore, in case of a discontinuous (lumpy) action space, like the binary option of cooperation or defection in a prisoner's dilemma game, there will be a *tipping point* for this drift to have a behavioral effect. For example, if the payoffs in a prisoner's dilemma game are, respectively, 0 (sucker), 1 (defection), 2 (cooperation), and 3 (temptation), then a participant will switch from defection to cooperation once her or his tie value or GTV would grow larger than ½. Ties developed in a related context may serve as stepping stone (eqs. (5b) and (6b) of the *i*ATM model).

(3) *Endogenous learning rate*: *short-term versus long-term*

In the *i*ATM model an affective tie is built up by the cumulative impact of impulses – that is, the experienced beneficial or detrimental behavior of another agent – weighted by a factor (δ_{ijt}) related to the extent of type uncertainty relative to behavioral uncertainty (see eqs. (2) and (4)). In a stationary environment the type uncertainty would be gradually resolved by repeated interactions (cf. eq. (3)), making δ_{ijt} an endogenous, gradually decreasing learning rate. Using eqs. (3) and (4), and applying the principle of insufficient reason ($\sigma_{ijt}^2 = \sigma_\varepsilon^2$), it can be calculated that it would take 9 interactions to get from $\delta_{ijt} = 0.5$ to $\delta_{ijt} = 0.1$, and 19 interactions to get to $\delta_{ijt} = 0.05$. Alternatively, and interestingly, if type uncertainty is much larger than behavioral uncertainty ($\sigma_{ijt}^2 \gg \sigma_\varepsilon^2$) it would take only about one interaction more to get to these same values from, in that case, $\delta_{ijt} \to 1$. Thus, the actions of the other agent will gradually lose impact, as no further information can be extracted from them. This may explain why some responses seem conditional whereas other responses do not (cf. Massen et al. 2010), even though both kinds of responses are (ultimately) based on the performance of the other agent. It is also consistent with findings suggesting less behavioral and neural impact of another individual's behavior once the other is perceived as being either good or bad (Delgado et al. 2005).

However, this holds in a stationary environment. In case of volatility, that is, repeated random shocks to the type distribution (see eqs. (3a) and (4a)), agents will start to rely more on impulses (with a larger learning rate δ_{ijt}). Consequently, they will show more impulsive behavior. For that reason a community facing an influx of immigrants might become less tolerant. On the other hand, recall from the previous section that in case of a threatening environment the impact of non-negative interaction experiences on affective tie formation may grow larger, if they become more surprising.

The *i*ATM model relies on a smart (Bayesian) emotional bookkeeping system, involving neural pathways honed through evolution, which allow the integration of episodic memories across time and contexts. As a consequence, refined responses in an interaction are facilitated. Nevertheless, simpler TFT-like responses can be expected in case of novel encounters or in a very volatile environment. A similar expectation holds for relatively simpler species lacking the more sophisticated emotional bookkeeping capacity.

(4) *Context, attention, and superposition*

Contexts play an important role in the *i*ATM model. Substantial evidence shows that behavior can be influenced by the way a situation is assessed. This has been observed, for example, in many so-called framing experiments (Tversky and Kahneman 1981, Kahneman 2011, De Martino et al. 2006). Simply changing the name of the game from Wall Street Game to Community Game can make a difference (Liberman et al. 2004). A context concerns an interaction episode generating an episodic memory. An open question is how contexts over time are represented in memory. Studies suggest that such episodes may be remembered seperately or can get integrated (Shohamy and Wagner 2008, Gershman and Daw 2017). The latter would seem to require less memory capacity and less effort for retrieval. The neurobiological support for friendships suggest that this may happen in case of tie formation (see Section 3). Although, even in that case, a distinction between (higher-level) contexts might be relevant. For example, a friend may be considered more or less “friendly” depending on the context attracting attention. In the literature on cooperation the pre-eminent context focused on is the prisoner’s dilemma (or related games). But, other likely relevant (modern) contexts are competitive markets and hierarchical settings such as states or firms. It seems plausible that in tie formation interaction episodes related to behavioral experiences in at least these three archetypical contexts – with a horizontal (competitive-cooperative) relationship dimension and a vertical (subordinate - dominant) relationship dimension – get associated in memory (either seperately or in some integrated form).

The *attentional weight* attached to a particular context in the decision-making process (see Section 2) is likely to be influenced by such factors as, on the one hand, its timing, the hedonic value or experienced utility of the interaction experienced within that context, and its similarity with the decision context at hand, and, on the other hand, the available mental resources. Consequently, a counterpart’s tie value may lose decisional impact over time not so much because of decay of the tie value as such but because relatively greater and thus more costly effort may be required to retrieve the estimate.

Apart from the effect shown by framing experiments that behavior can be influenced by manipulating the assessed context, an important consequence of this contextual view is another “mind the gap” caveat. This time it is not to point at a behavioral difference between described uncertainty and experienced uncertainty (Hertwig and Erev 2009) but between behavior in an unspecified context versus a specific context. For instance, if asked in a survey whether one would trust an anonymous randomly selected stranger, one will have to answer in a kind of *superposition* regarding contexts. The answer provided is likely to differ from the (behavioral) response given once the interaction context with that stranger is known (compare the GTV specifications of eqs. (6a) and (6b) of the *i*ATM model). This may help explain why there is no conclusive evidence regarding the correlation between responses to the World Values Survey Trust Question and experimental trust game findings (see Banerjee 2016). Finally, from a contextual point of view, a difference exists between helping in a dictator game with no possibility of a response affecting the dictator and helping an anonymous stranger in the street, where some response cannot be excluded. This would seem to make the social value orientation task referred to in the previous section – which involves someone else simultaneously making a decision affecting the performer of the task – an ecologically more relevant measure of helping than a dictator game.

(5) *Generalizations*: *propagation of caring, bonding under nonrecognition, and personification*

The GTV construct of the *i*ATM model (see eqs. (6)-(6b)) fosters a *propagation* of caring induced by interaction experiences, over time (persistence) as well as across individuals (spreading). Notably, the model fairly accurately predicts the “cooperative behavior cascades” findings of Fowler and Christakis (2010). Using data from Fehr and Gaechter (2002) regarding repeated one-shot public good game experiments with groups of four participants, they find that the influence of a participant's contribution behavior persists for multiple periods and spreads up to several degrees of separation (across individuals). More specifically and to begin with the latter, their results show that for each monetary unit contributed by an alter, one period back, ego contributes an additional 0.19 units. For each unit contributed by alter's alter, two periods back, ego contributes an additional 0.07 units, and for each unit by alter's alter's alter, three periods back, 0.03 units (albeit significant at the 20% level only). To arrive at related model predictions the following assumptions are made for the *i*ATM model: $\sigma_{ijt}^2 = \sigma_\varepsilon^2$ (applying the principle of insufficient reason), type-risk neutrality, a reference contribution of 0, and a smooth utility function such that the normalized contribution level is determined by the tie (cf. impulse equation above eq. (1)). Then, a simple calculation shows that the *i*ATM model predicts the following additional contribution as fraction of the initial contribution (impulse) for period t = -1,-2,-3, indicating the relevant period in the past (the relevant alter): $(\delta_{ijt}/3)^{-t}$, with $\delta_{ijt} = 1/2$ for all t. Consequently, the corresponding predictions for a given impulse dating 1-3 periods back, respectively, equal: 0.17, 0.03, and 0.01. These predictions are quite similar to the empirical observations of Fowler and Christakis. The authors further find persistence effects in that an alter influences ego's behavior up to four periods later, with the extra amount per impulse unit successively being equal to: 0.19, 0.15, 0.08, and 0.17 (which seems more like an outlier, as the next amount is 0). Under the same assumptions as before, the *i*ATM model predicts the following amount for period t = 1,2,… : $(1-\delta_{ijt})^{t-1}(\delta_{ijt}/3)$, with again $\delta_{ijt} = 1/2$ for all t. The corresponding predictions are therefore: 0.17, 0.08, 0.04, and 0.02. Again, these predictions are remarkably close to the findings of the authors (apart from the 'outlier'). Through these channels substantial propagation of caring impulses may take place.

Another kind of generalization may occur under *nonrecognition*, that is, if individual agents cannot be recognized, either due to the absence of detection or memory skills (such as with relatively more primitive organisms) or because of total anonymity (making identification impossible). As mentioned in Section 2, it would be like interacting with a single agent, even though the actions may stem from different agents. Although the same model specification would hold as for a single agent, the stochastics of the environment would be affected, with more volatility to be expected if different agent types are present. In that event the *i*ATM model would predict greater reliance on impulses (a larger learning rate) and thereby more TFT-like behavior. Biological studies concerned with this type of situation speak of "generalized reciprocity" by animals, where the decision to cooperate is based on the outcome of the last encounter even if with a different partner, and of "quorum sensing" by bacteria, where cell activity is triggered if sufficient activity by neighboring cells is detected (see, e.g., Pfeiffer et al. 2005, Allen et al. 2016). For modern humans, this type of generalization may hold for anonymous web platforms, like Reddit, where users anonymously contribute to a public (club) good.

A further generalization concerns what seems to resemble the *personification* or animation of objects ("object love"). This can lead to what may be interpreted as caring, such as when participants in experiments costly retaliate to random choices by nature (Blount 1995) or, while being aware that the other players in the game are computers, play cooperatively in a public good game (Houser and Kurzban 2002). Furthermore, evidence suggest that attachment may be important in explaining the endowment effect, where people ask more money for giving up a good than they are willing to pay for acquiring it (Ariely et al. 2005). Such affective tendencies may be controlled by higher-order cognitive functions but only with some effort. An important application in this context, exploiting the existence of an affective tie mechanism, concerns the development of eldercare robots as interaction companion and helper.

(6) *Affective networks and collective action in small and large groups*

Evolutionary biologists have suggested that people can maintain at most around 150 stable social relationships (Dunbar 1998). This upper-limit appears to cover the sizes of hunter-gatherer groups that our forebears lived in for most of the time in evolution. In such a setting the affective tie mechanism helps explain why (some) groups succeeded in overcoming the public good (free-rider) problem, including the organization of defense or offense towards out-groups, which in turn may have played a role in the evolutionary selection of the mechanism (see, e.g., Bowles and Gintis 2011).

Its relevance extends in time to the functioning of modern local communities and organizations (van Dijk and van Winden 1997). In the political sphere, *affective networks* provide a new angle for understanding political participation, such as via voting, lobbying, and resistance (van Winden 2015). Affective ties with policymakers help to get access and influence, while networks of friends are instrumental in grassroots movements – like political campaigns and demonstrations – facilitated by social media. Even (non-atomistic) markets may be better understood this way. A case in point concerns spontaneous market collusion, without any document being necessary, among officials of firms with ties based on frequent interaction (van Winden 2012).

Affective ties, furthermore, play a crucial role in the instilling and internalization of social *norms*, important to the functioning of groups. The reason is that these social psychological processes run on the software of the mechanism. To the extent that positive ties exist with caretakers and educators positive emotions will be triggered by their approval of good behavior and negative emotions by their disapproval of bad behavior, which over time leads to the (anticipated) experience of shame or pride when the norm is violated or adhered to, respectively. This way the norm gets internalized and its satisfaction becomes an intrinsic motivation. Note, however, that here the intrinsic motivation is future-self oriented and not related to an other-directed caring social preference (see item (8) for more on temporal selves).

With larger groupings like modern societies an additional type of dyadic relationship exploiting the affective tie mechanism becomes important, which is *emotional leadership*, where individual followers bond with charismatic leaders (van Winden 2015, Loerakker and van Winden 2017). History provides many examples of charismatic leaders exploiting such relationships to produce large scale public goods (or bads), because they can count on contributions by followers. One may think of Gandhi, Mandela, King or Hitler. Moreover, these ties may further cause followers to help each other, either to please the leader or because direct ties develop between them, such as through interaction at gatherings with the leader. An interesting extension in this context concerns Lijphart's theory of pacification in a pillarized society of diverse cohesive groups through bridging leaders (elite cartels), which in his view characterized the Netherlands between 1917 and 1967 (Lijphart 1975). In case of hardly interacting members of groups with different ideological, ethnic or religious backgrounds, this may be an optimal network architecture. The downside may be that this lack of interaction is reinforced by this architecture.

(7) *In-group versus out-group, and the role of threat*

Group formation may readily happen under the influence of the affective tie mechanism. A special role in this respect is played by the interaction between parents and infants (see Section 3). Through the spread and persistence of affective ties (see item (5) above) this has a much wider effect on social interactions. Especially, as evidence indicates that the quality of the attachment to (parental) caretakers has a long-term effect on the neural circuitry for the mechanism in the child, influencing its capacity to make affective bonds later in life (Newman et al. 2015, Numan 2015). Recall, moreover, that showing care for someone else may induce care and proximity seeking in reverse as well. Because of the concomitant effects of positive tie formation, such as attention and proximity seeking, positive affective networks will easily manifest themselves as in-groups, with in-group love and trust.

Abundant evidence further shows that bonding and prosociality may increase under *external threat*, like war violence, with persistent effects that are related to changes in preferences (Gneezy and Fessler 2012). Because nonnegative impulses from group members may get greater weight in a threatening situation (see Section 3), and positive ties with group members greater attention (buffering effect), the *i*ATM model helps explain. Importantly, the *i*ATM model does *not* predict, however, that in-group love implies out-group hate. This, in contrast to the nature of the "parochial altruist", defined to be driven not only by generosity toward fellow group members but also hostility toward outsiders, that is central to some evolutionary explanations of cooperation (Bowles and Gintis 2011). Consistent evidence comes from intergroup prisoner's dilemma experiments focusing on the impact of social value orientation (related to GTV in the model) where, compared to proselfs, prosocials are found to display stronger in-group trust and in-group love – that is, they contribute more to benefit their in-group – but no different out-group distrust and out-group hate; moreover, the effect on in-group help appears to be mediated by in-group trust (De Dreu 2010; see also Halevy et al. 2008, Weisel and Böhm 2015). Negative ties and related emotions and preferences would only emerge through negatively appraised interaction, such as via directly experienced aggression or through the observed suggestive behavior of other in-group members (which can activate ties or instilled norms; cf. Bauer et al. 2016, and item (6) on norms). Without any prior interaction, one would rather expect an emotion like wonder, curiosity or excitement on meeting a stranger.

When the behavior of strangers is appraised as threatening, negative impulses feed the *i*ATM, thereby generating an intrinsic motivation to avoid or to harm. This may happen, for instance, if people (like immigrants) from different ethnic groups become associated with violence, terrorism or viruses. As, in a more threatening environment, greater variance (volatility) and skewness to the right of the type probability distribution seem likely, the *i*ATM model would predict a larger reliance on impulses (less tolerance) together with a relatively larger impact of (more surprising) positive experiences. The latter requires interaction, however, which would become more difficult if negative ties are formed. One route would be via bridging friendships, as discussed in item (2). Another potential route concerns participation in and thereby contributing to joint positive experiences. For example, Depretis-Chauvin et al. (2020) find that shared experiences with victories of national football teams in Africa boost national identity, while diminishing identification with one's own ethnic group. It also increases trust and decreases conflict across ethnic groups, without being matched by an enhanced dislike for foreigners. (For more on identity, see item (9) below.) Incidentally, the widely implemented policy of social distancing and isolation in the Covid-19 pandemic, where others are threatening, is clearly problematic in this respect, because it works against contact and the exploitation of its buffering effect.

(8) *Social preferences, risk preferences, and time preferences linked through uncertainty*

The *i*ATM model focuses on social uncertainty and related social preferences due to affective ties. Preferences usually related to uncertainty in economics are risk preferences, concerning the outcome of a lottery or an investment, and time preferences regarding future events. For all preferences, the standard assumption holds that they are stable and given (for exceptions, see Becker and Mulligan 1997, Bowles 1998, Doepke and Zilibotti 2017). In stark contrast, it will be argued here that the affective tie mechanism not only influences social preferences but may also influence both risk and time preferences, making the latter dependent on interaction experiences as well.

To start with *risk preferences*, it is important to distinguish nonsocial from social risks, that is, whether the outcome of one's choice is determined by a lottery (like a dice) or another individual (as in a trust game). In the former case, the *i*ATM model suggests that the decision will be determined by the extended utility (eq.5(b)) of the decision-maker, where the extension is related to the affective ties attended to in the given context. For instance, positive affective ties (attachment) with parental caretakers, attended to in a risky decision-making context, may make offspring more risk-averse if parents are perceived to prefer risk-aversion (or, alternatively, have instilled risk aversion as a personal norm). This may help explain the finding of Falk and Kosse (2016) for preschool children and young adults that having experienced a longer duration of breastfeeding (a valid measure of attachment, see Gibbs et al. 2018) is associated with greater risk aversion. The lack of a barrier of letting parental caretakers down (attachment) or oneself (internalized norm) may also help explain the finding of Hoeve et al. (2012) of poor attachment being correlated with juvenile delinquency.

Regarding social risk, an interesting case in point concerns the trust game, where the trustor can decide to transfer any amount T (≥ 0) out of a given monetary endowment to a trustee, which is then (typically) tripled and received by the trustee, who can in turn back-transfer any amount (≤ 3T). From the perspective of the *i*ATM model, transfering money in this game is very different from choosing a lottery, because now another agent (the trustee) is involved, which makes the trustor's decision dependent on both her or his extended utility (eq. 5(b)) and GTV (eq.(6b)). Experimental results show that on average transfers equal about 50% of the trustor's money, while back-transfer are more or less equal to the transfer T (Glaeser et al. 2000, Camerer 2003, meta-analysis by Johnson and Mislin 2011).

Although the trustor's behavior is unlikely to be fully explained with the *i*ATM model, as the trustor's GTV is likely to be small and the trustor is incentivized to think ahead, which stimulates self-control, two pieces of evidence seem worth mentioning here. First, Kosfeld et al. (2005) find in a risk experiment where the investor faced the same choices as in the trust game but in which a random mechanism (instead of a trustee) determined the outcome – and, thus, investors faced exactly the same amount of risk as in the trust game – that the average transfer was about the same. This is puzzling in light of expected betrayal aversion in the trust game (Bohnet and Zeckhauser 2004, Bohnet et al. 2008). However, the trustor's GTV may sufficiently counterbalance betrayal aversion and may thereby help explain this result. This is supported by a second piece of evidence, regarding a positive relation between social value orientation and cooperation in social dilemmas (Cornelissen et al. 2011), which is consistent with an expected positive impact of a trustor's GTV. The reason is that a trust game can be seen as a sequential prisoner's dilemma, with (back-)transfering money seen as cooperation and not (back-)transfering money as defection (Camerer 2003).

The *i*ATM model should be of greater relevance for explaining the trustee's response, due to the likely more emotional state of the trustee after receiving the transfer. Moreover, in comparison with the trustor, for the trustee's behavior the impact of the transfer as impulse will play an additional role, as it affects jointly with the trustee's GTV as prior the tie with the trustor (see eqs. (2) and (4) of the model). For illustration, let T_r (T_e) stands for the trustor's (trustee's) transfer and Z for their endowment, then the finding $T_e = T_r$ for $T_r/Z = ½$, can be predicted by the *i*ATM model, assuming log-utility (V_e) for the trustee, that is: $V_e = ln(Z + 3T_r - T_e) + \alpha_{er} ln(Z - T_r + T_e)$. As also the trustee's GTV is likely to be small, and the relevant reference transfer may be zero, the only requirement is that the weight attached to the impulse (δ) is large (goes to unity), demanding that type uncertainty is much larger than behavioral uncertainty. (Incidentally, note that $T_r/Z = ½$ is compatible with the maximization of the trustor´s utility $ln(Z - T_r + T_e)$ if $T_e = T_r$.)

The potential relevance of the *i*ATM model regarding *time preferences* is suggested by some recent studies in (neuro)psychology distinguishing between temporal selves, in particular, a future self (FS) and a present self. These studies provide supportive behavioral as well as neural evidence for the view that the FS is perceived as another person with whom similar affective relationships (psychological closeness or connectedness) can be developed as with other people (see also the next item, on identity). First, experimental evidence shows that decisions for FS, like savings, are similar to decisions for other people (Pronin et al. 2008, Bartels and Rips 2010). Second, brain activity triggered by thinking about FS shows a similar pattern of activation as thinking about another person (Hershfield 2011, Soutschek et al. 2016). Soutschek et al. find that a particular brain region (TPJ, see Section 3) is involved in both future-oriented behavior and in overcoming egocentricity bias in social discounting (that is, caring less for mentally more distant others). In both cases, attention appears to be shifted away from the perspective of the present self. While current theories of intertemporal decision-making focus on the brain's top-down planning functions in the PFC (McClure et al. 2004, Figner et al. 2010), these findings suggest an alternative, more bottom-up route for patience which involves care for others (including FS). In Section 3, the same brain region (TPJ) was found implicated in the encoding of affective ties. Following the *i*ATM model, a positive (or negative) attitude towards FS may be associated with an individual's GTV determined by the affective ties attended to in a context of intertemporal decision-making. Early in life, before any or much experience with such decisions, these ties are likely to be more indirect, produced by the attachment to caretakers (parents, educators), through whose attitude patience may or may not be stimulated. Suggestive evidence comes from the experimental findings of Falk and Kosse (2016), regarding both preschool children and young adults, that having experienced a longer duration of breastfeeding (see above) is not only associated with greater risk aversion but also with greater patience; in both datasets they control for factors like cognitive ability and socio-economic status. Once experience accumulates with the performance of one's FS – which may trigger disappointment or joy and gratitude (DeSteno et al. 2014) –, a direct tie with FS is likely to develop, with a more direct influence on intertemporal discounting. Interestingly, individuals with an autism spectrum disorder face not only difficulty with social bonding but also with mentally reexperiencing past states (emotional episodes) and preexperiencing future states, essential for grasping the self's continuity through time (Lind 2010, Lind and Bowler 2010). The role of affect is further demonstrated by the impact of emotional empathy in both intertemporal decision-making and care for others (Decety and Svetlova 2012, O'Conell et al. 2013), and of emotional intensity factors like vividness, realness, and proximity (Loewenstein 1996, Palombo et al. 2015, Alan and Ertac 2018).

(9) *Happiness and identity*

There is a notable shift in economics from the standard material welfare concept, where utility is defined in terms of goods and services or money, towards a concept of welfare related to *happiness* (Oswald 1997, Frey and Stutzer 2002, Clark et al. 2018). In fact, this involves a partial return to scholars like Jeremy Bentham, who focused on pleasure and pain (Bentham 1789). A robust empirical finding reported by happiness studies is that social contacts with beloved ones and friends are a major source of happiness. Also, through their impact on mental health – in itself a key predictor of happiness –, with the quality of social relationships counting as a major risk factor for depression (Clark et al. 2018, Teo et al. 2013; see also Cacioppo et al. 2009 for references to more general health problems due to social isolation). The *i*ATM model provides a tool for understanding the nature of this source of happiness, and how it can be developed or maintained. According to this uncertainty-based model affective ties with other people – or personified objects, for that matter – and their welfare are a key source of (extended) utility; that is, to the extent these ties are attended to in a given context. An individual's extended utility reflects whether the environment is friendly – providing opportunities – or hostile, and can therefore be trusted (approached) or should be distrusted (avoided or fought), respectively. Regarding attention, this would suggest, for example, that self-focused attention in a mental depression (Brockmeyer et al. 2015) may thwart recovery by blocking mental access to one's affective network.

In the *i*ATM model, past beneficial interaction experiences motivate helping others in order to increase one's extended utility (see eqs. (5) and (7)). Note that this prediction runs counter to studies suggesting that just spending money on others would promote happiness (Dunn et al. 2008). However, the findings of Dunn et al. are either based on correlations (regressions) or on an instruction to spend the money just received in the experiment "on a gift for someone else or charitable donation" before the end of the afternoon (which was then followed by a second, ex-post happiness reporting). Although interpreted as causal, the model suggests that this procedure may have induced participants to pay attention to their affective ties, activating their extended utility, which in turn caused their spending to be associated with a higher happiness rating. Consistent with this view, Vohs et al. (2006) find that priming participants with money (activating the idea of an abundance of money) induces a self-sufficient orientation and reduces helpfulness towards others.

The *i*ATM model has a bearing also on the formal modeling of *identity*. In an influential paper, Akerlof and Kranton (2000) see identity as "a person's sense of self", where they focus on the role of social norms. They propose to include identity into the utility function, where "identity is associated with different social categories and how people in these categories should behave". Their proposal concerns a sociopsychological extension of the standard (narrowly selfish) utility function in economics. The *i*ATM model suggests a biopsychological extension of the standard approach focusing on the role of affective ties for a person's "sense of self". One's identity is here proposed to comprise all agents – selves and others – whose utility one is willing to work for (albeit to a different, attention and tie-value related extent). Consistent with this proposal, for example, is Hershfield's argument regarding the effort of long-term planning that what matters is an identity comprising both the current and the future self, although there can be degrees of psychological connectedness (Hershfield 2011; see also Parfit 1971). Similarly, the intrinsic motivation to live up to certain norms is related to the connectedness with one's future self (see item (6) above).

Finally, the *i*ATM model's dynamic view of identity provides a link with the biological concept of organismality, as discussed in Section 3. Agents with strong positive ties may appear to behave like one organism. A family household is an interesting case in point. Traditionally taken as a single agent in economics, in modern labor economics it is rather seen as a group of family members that exchange and bargain among each other. The traditional single-agent assumption would be valid, however, if mutually strong positive bonds between family members may be expected; particularly, if tie values towards other household members are equal or close to unity, making each member interested in promoting their joint household welfare (Becker 1974).

(10) *Cooperation, competition, hierarchy, and the French revolutionary motto*

In item (4) above two relationship dimensions were distinguished: one horizontal running from extremely competitive to extremely cooperative, and one vertical running from extremely subordinate to extremely dominant. By relating competitiveness to conflict and antipathy, and cooperation to sympathy, these dimensions can be linked to the *i*ATM model by having the horizontal dimension represented by the tie value ($-\infty < \alpha_{ijt} < +\infty$), and the vertical dimension by control over choice sets (ranging from no control at all – leaving an empty choice set – to full control over one's own and the other's choice set). Greater control means less uncertainty and a more 'friendly' and manipulable environment. This may be represented by a more negatively skewed type distribution for the controlling dominant agent, making positive (negative) actions from the controlled subordinate agent less (more) surprising, whereas the reverse may hold for the subordinate agent. Consequently, the tie value of the more dominant agent may be lower than the tie value of the more subordinate agent, and the same holds for affective empathy, as tie value and affective empathy appear to be related (Batson et al. 2007; see also Section 3). In turn, this may lead to less affective and relatively more cognitive empathic behavior (such as in manipulation) by the dominant agent. In this context, recall the discussion of exploitative leaders in item (6). Incidentally, this may help explain the findings of Piff et al. (2010) that higher (lower) class / socioeconomic status individuals are less (more) helpful, generous and trusting, and show more (less) unethical behavior (Piff et al. 2012).

The three archetypical contexts that were related to these two dimensions in item (4) – hierarchy, competition, and cooperation – can be associated in that order with the following three coordination mechanisms distinguished in the literature on the governance of scarce resources: government (coercive power), market, and community (see, e.g., Ostrom 1990, Rajan 2019). With one notable caveat concerning competition. In economics, extreme competition in markets, called perfect or atomistic competition, is characterized by the absence of any individual behavioral influence (agents are "price takers" and can freely adjust their demand or supply without directly affecting others). In this ideal-typical context, the affective tie mechanism will have no bite (nor has domination), which puts it at the crossing of the two dimensions, and not at the extreme competitive relationship side (where "cut-throat competition" would better fit). Only with nonatomistic competition and the relevance of exchange relationships, due to behavioral impact, competition in the here maintained sense occurs. Modern economics also deals with such impactful behavior in markets, using game theory as an important tool, while political economics focuses on the behavioral relationships of governmental agents (like that of politicians, bureaucrats, and voters). The role of communities in the allocation of resources – that is, by affective networks of interpersonal relationships – has been neglected, though (see Rajan 2019). Nobel laureate 2009 Elinor Ostrom has played an important role in bringing community like relationships to the fore in the context of cooperation in the use of common pool resources (Ostrom 1990, 2010, 2012; Ostrom et al. 1994). As with the provision of a public good, this use is characterized by behavioral externalities, that is, welfare effects on others that are not accounted for in narrowly selfish decision-making, leading to an inefficient allocation of resources (a social welfare loss). Ostrom challenged the conventional wisdom – based on the folk psychological model of a fully rational and selfish agent (*homo economicus*) – that such resources should be either regulated by a central authority or privatized. Her challenge was based on evidence from field as well as lab studies showing that voluntary sanctioning related to social norms and reciprocity like behavior may attenuate the neglect of others. The *i*ATM model brings in another, affect-based, facilitator, which is that people may have an intrinsic motivation to care for each other and have a concern for each other's welfare (van Winden 2012, see also Reiss et al. 2000). According to Ostrom (2012, p60): "The possibility that there are individuals who take into account the payoffs of other individuals changes theoretical foundations greatly."

Item (6) discussed the importance of affective networks for collective action, such as for the provision of local public goods or political action. The nature of the affective tie mechanism entails dynamic linkages between the functioning of a governments, markets, and communities. For example, affective ties may be crowded out – with a negative effect on the social cohesion and related welfare of communities – by the encroachment of the government or the market on the local provision of goods, making social interaction required for the provision of these goods superfluous (van Dijk and van Winden 1997, Rajan 2019). For another example, if communities and thereby affective networks wither away under the encroachment of markets (or the government), it may affect political participation by citizens and consequently the redistribution of resources and the regulation of markets through a representative government.

At a deeper practical philosophical level this hints at the importance of the third element of the French revolutionary motto: *liberté, égalité, fraternité.* In the wake of the enlightment and the development of liberal market democracies liberty and equality got emphasized, relative to fraternity, which arguably thwarted an optimal balance between the three coordination mechanisms (Rajan 2019). This is due to the aforementioned role of affective networks in overcoming collective action (free-riding) problems necessary for attaining equality in political participation (including, but not restricted to voting). Political equality in turn facilitates liberty and equality in the private sector through the safeguarding of security and the enforcement of property rights, an appropriate redistribution of resources, and adequate regulation of markets. Moreover, political participation facilitates the transmission of information regarding the interests of citizens necessary for a representative government.

5 Conclusion

A notable similarity across species is that organisms are confronted with deep environmental uncertainty regarding opportunities and threats, in particular, regarding the friend-or-foe nature of their counterparts in interactions. For their survival and flourishing some mechanism helping them to detect the friend-or-foe type of other agents and motivating them to seek their proximity if assessed as beneficial, or to avoid (or fight) them if assessed as harmful, would seem to be essential. The informational affective tie mechanism of the here proposed *i*ATM model serves this twofold requirement. The support for this model summarized in this paper, coming from very different data sources and fields of research, is substantial. Furthermore, the many discussed implications and applications may testify to the wide scope and relevance of the model and its potential for organizing findings and generating testable hypotheses in these different fields. Further modeling and testing is needed. This particularly holds for the relationship between the attentional weight attached to a context and its driving factors, in particular, the timing of the context (interaction episode), the hedonic value of the experienced interaction, and the similarity with the context at hand (see Bordalo et al. 2020 for an interesting related modeling attempt concerning consumption option choices; see also Tversky 1977, Gilboa and Schmeidler 1995, Mullainathan 2002, Gershman and Daw 2017). Moreover, further testing is needed with respect to the robustness of the findings presented, new hypotheses and predictions that may be derived, and new applications of the model which may fruitfully involve species other than humans.

Two final remarks in concluding. First, recall that the automatic informational affective tie mechanism proposed and studied in this paper need not always have positive social effects, because it may generate hatred through negative affective ties. Also, affective networks (in-groups, communities) may develop and maintain negative affective ties with other such networks and may stand in the way of creativity and innovation, as new ways of behaving or thinking may not be liked by group mates. In the former event, bridging positive ties (requiring the opportunity of interaction) may contribute in tearing down affective and cognitive stereotyping barriers, while in the latter also competition between groups may help overcome resistance. Second, although no longer the exogenously given preferences of standard economic theory, the endogenous social preferences predicted by the *i*ATM model are systematic, persistent, and measurable. Nevertheless, the dynamics (hysteresis) as well as context- and attention-dependency predicted by the model may help explain why the robustness and prediction reliability of many behavioral and neural findings have come into question lately.

References

Abraham, E., Hendler, T., Shapira-Lichter, I., Kanat-Maymon, Y., Zagoory-Sharon, O., and Feldman, R. (2014). Father's brain is sensitive to childcare experiences. *PNAS*, *111(2)*, 279792–9797.

Ackermann, K. A., and Murphy, R. O. (2019). Explaining cooperative behavior in public goods games: How preferences and beliefs affect contribution levels, *Games*, *10*, 15, 1-32.

Akerlof, G. A. (1983). Loyalty filters. *American Economic Review*, *73*, 54–63.

Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, *115(3)*, 715-753.

Alan, S., and Ertac, A. (2018) Fostering patience in the classroom: results from randomized educational intervention. *Journal of Political Economy*, *126(5)*, 1865-1911.

Allen, R. C., McNally, L., Popat, R., and Brown, S. P. (2016). Quorum sensing protects bacterial co-operation from exploitation by cheats. *ISME Journal*, *10*, 1706–1716.

Ariely, D., Huber, J., and Wertenbroch, K. (2005). When do losses loom larger than gains? *Journal of Marketing Research*, *42(2)*, 134-138.

Aureli, F., and Schaffner, C. M. (2002). Relationship assessment through emotional mediation. *Behaviour*, *139*, 393-420.

Aureli, F., and Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. *Behavioral Ecology and Sociobiology*, *73*, 6.

Axelrod, R. (1980). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24(3)*, 379-403.

Axelrod, R., and Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211(4489)*, 1390-1396.

Báez-Mendoza, R., and Schultz, W. (2013). The role of the striatum in social behavior. *Frontiers in Neuroscience*, *7(233)*, 1-14.

Balliet, D. P., Parks, C., and Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Processes and Intergroup Relations*, *12(4)*, 533-547.

Baluška, F., Volkmann, D., Hlavacka, A., Mancuso, S., and Barlow, P. W. (2006). Neurobiological view of plants and their body plan. In Baluška, F., Mancuso, S., and Volkmann, D. (Eds.), *Communication in Plants*. Berlin: Springer-Verlag.

Banerjee, R. (2016). On the interpretation of World Values Survey trust question: global expectations vs. local beliefs. *IZA* Discussion Paper, No. 9872.

Bartels, D. M., and Rips, L. J. (2010). Psychological connectedness and intertemporal choice. *Journal of experimental Psychology: General*, *139(1)*, 49–69.

Bartz, J. A., Zaki, J., Ochsner, K. N., Bolger, N., Kolevzon, A., Ludwig, N., and Lydon, J. E. (2010). Effects of oxytocin on recollections of maternal care and closeness. *PNAS*, *107(50)*, 21371-21375.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., and Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality and Social Psychology*, *93*, 65-74.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., and Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, *30(3)*, 249–274.

Bault, N., Fahrenfort, J. J., Pelloux, B., Ridderinkhof, K. R., and van Winden, F. (2017). An affective social tie mechanism: Theory, evidence, and implications. *Journal of Economic Psychology*, *61*, 152–175.

Bault, N., Pelloux, B., Fahrenfort, J., Ridderinkhof, K. R., and van Winden, F. (2015). Neural dynamics of social tie formation in economic decision-making. *Social Cognitive and Affective Neuroscience*, *10*, 877–884.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, *82(6)*, 1063-1093.

Becker, G. S., and Mulligan, C. B. (1997). The endogenous determination of time preference. *Quarterly Journal of Economics*, 729-758.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*, 1652-1678.

Ben-Ami Bartal, I., Rodgers, D. A., Sol Bernardez Sarria, M., Decety, J., and Mason, P. (2014). Pro-social behavior in rats is modulated by social experience. *eLife*, 3, e01385.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. New York: Hafner (1948).

Bethlehem, R. A. I., van Honk, J., Auyeung, B., and Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies. *Psychoneuroendocrinology*, *38*, 962—974.

Bickart, K. C., Dickerson, B. C. and Feldman Barrett, L. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, *63*, 235–248.

Blount, S. (1995). When outcomes aren't fair: the effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, *63(2)*, 131-144.

Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: evidence from Brazil, China, Oman, Switzerland, Turkey and the United States. *American Economic Review*, *98(1)*, 294-310.

Bohnet, I., and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, *55*, 467–484.

Bogaert, S., Boone, C., and Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*, 453–480.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*, *135(3)*, 1399–1442.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, *36*, 75–111.

Bowles, S., and Gintis, H. (2011). *A Cooperative Species.* Princeton: Princeton University Press.

Brandts, J., Riedl, A., and van Winden, F. (2009). Competitive rivalry, social disposition, and subjective well-being: An experiment. *Journal of Public Economics*, *93*, 1158–1167.

Brenner, E. D., Stahlberg, R., Mancuso, S., Vivanco, J., Baluška, F., and Van Volkenburgh, E. (2006). Plant neurobiology: an integrated view of plant signaling. *TRENDS in Plant Science*, *11(8)*, 413-419.

Brockmeyer, T., Zimmermann, J., Kulessa, D., Hautzinger, M., Bents, H., Friederich, H-C., Herzog, W., and Backenstrass, M. (2015). Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in Psychology*, *6(1564)*, 1-10.

Brosnan, S. F., and de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*, 129–152.

Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, *6(147)*, 1-19.

Brucks, D., and von Bayern, A. M. P. (2020). Parrots voluntarily help each other to obtain food rewards. *Current Biology*, *30*, 1–6.

Cacioppo, J. T., Fowler, J. H., and Christakis, N. A. (2009). Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology*, *97(6)*, 977–991.

Camerer, C. F. (2003). *Behavioral game theory*. Princeton: Princeton University Press.

Carter, G. G., Farine, D. R., Crisp, R. J., Vrtilek, J. K., Ripperger, S. P., and Page, R. A. (2020). Development of new food-sharing relationships in vampire bats. *Current Biology*, *30*, 1–5.

Cheng, Y., Chen, C., Lin, C. P., Chou, K. H., and Decety, J. (2010). Love hurts: An fMRI study, *NeuroImage*, *51*, 923–929.

Churchland, P. S., and Winkielman, P. (2012). Modulating social behavior with oxytocin: How does it work? What does it mean? *Hormones and Behavior, 61*, 392–399.

Clark, A. E., Flèche, S., Layard, R., Powdthavee, N., and Ward, G. (2018). *The Origins of Happiness: The Science of Well-Being over the Life Course*. Princeton: Princeton University Press.

Coleman, J. S. (1984). Introducing social structure into economic analysis. *American Economic Review*, *Papers and Proceedings*, *74*, 84–88.

Cooper, D. J. and Kagel, J. H. (2016). Other-regarding preferences: A selective survey of experimental results. In J.H. Kagel and A. E. Roth (Eds.), *The Handbook of Experimental Economics*, *vol. 2*, 217–89. Princeton: Princeton University Press.

Cornelissen, G., Dewitte, S., and Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, *37*, 1080–1090.

Cronin, K.A. (2012). Prosocial behaviour in animals: the Influence of of social relationships. *Animal Behaviour*, *84*, 1085-1093.

Damasio, A. (2018). *The strange order of things*. New York: Vintage Books.

Daw, N. D. (2014). Advanced reinforcement learning. In P. W. Glimcher and E. Fehr (Eds.), *Neuroeconomics* (2nd ed.). Amsterdam: Elsevier.

Dayan, P., Kakade, S., and Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, Supplement, 1218-1223.

Decety, J. and Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience, 2*, 1–24.

Declerck, C. H., Boone, C., and Emonds, G. (2013). When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition*, *81*, 95–117.

De Dreu, C. K. W. (2010). Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Processes & Intergroup Relations*, *13(6)*, 701–713.

De Dreu, C. K. W. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Hormones and Behavior*, *61*, 419-428.

de Freslon, I., Peralta, J. M., Strappini, A. C., and Monti, G. (2020). Understanding allogrooming through a dynamic social network approach: an example in a group of dairy cows. *Frontiers in Veterinary Science*, *7(535)*, 1-12.

Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611-1618.

De Martino, B., Kumaran, D., Seymour, B., and Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*, 684-687.

Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building nations through shared experiences: evidence from African football. *American Economic Review*, *110(5)*, 1572–1602.

DeSteno, D., Li, Y., Dickens, L., and Lerner, J. S. (2014). Gratitude: a tool for reducing economic impatience. *Psychological Science*, *25(6)*, 1262–1267.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279–300.

Diaz-Muñoz, S. L., Boddy, A. M., Dantas, G., Waters, C. M., and Bronstein, J. L. (2016). Contextual organismality: Beyond pattern to process in the emergence of organisms. *Evolution*, *70(12)*, 2669–2677.

Doepke, M., and Zilibotti, F. (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, *85(5)*, 1331–1371.

Douglas, A. J., and Russell, J. A. (2001). Endogenous opioid regulation of OT and ACTH secretion during pregnancy and parturition. In J. A. Russell et al. (Eds.), *Progress in Brain Research*, *133*, 67-82.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, *6(5)*, 178-190.

Dunbar, R. I. M. (2010). The social role of touch in humans and primates: Behavioural function and neurobiological mechanisms. *Neuroscience and Biobehavioral Reviews*, *34*, 260–268.

Dunn, E. W., Aknin, L. B., and Norton, M. I. (2008). Spending money on others promotes happiness. *Science*, *319(5870)*, 1687-1688.

Engelmann, J. B., Meyer, F., Ruff, C. C., and Fehr, E. (2019). The neural circuitry of affect-induced distortions of trust. *Science Advances*, *5* (eaau3413), 1-16.

Evers, E., de Vries, H., Spruijt, B. M., and Sterck, E. H. M. (2015). Emotional bookkeeping and high partner selectivity are necessary for the emergence of partner-specific reciprocal affiliation in an agent-based model of primate groups. *PLoS ONE*, *10(3)*, e0118921, 1-33.

Evers, E., de Vries, H., Spruijt, B. M., and Sterck, E. H. M. (2016). Intermediate-term emotional bookkeeping is necessary for long-term reciprocal grooming partner preferences in an agent-based model of macaque groups. *PeerJ*, *4*, e1488, 1-38.

Fahrenfort, J. J., van Winden, F., Pelloux, B., Stallen, M., and Ridderinkhof, K. R. (2012). Neural correlates of dynamically evolving interpersonal ties predict prosocial behavior. *Frontiers in Neuroscience*, *6(28)*.

Falk, A., and Kosse, F. (2016). Early childhood environment, breastfeeding and the formation of preferences. SOEPpapers 882.

Faraji, M., Preuschoff, K., and Gerstner, W. (2018). Balancing new against old information: the role of surprise in learning. *Neural Computation*, *30(1)*, 34-83.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*, 159-181.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137-140.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of*

Economics, *114*, 817–868.

Feldman, R. (2016). The neurobiology of mammalian parenting and the biosocial context of human caregiving. *Hormones and Behavior*, *77*, 3–17.

Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, *21(2)*, 80-99.

Fellbaum, C. R., Gachomo, E. W., Beesetty, Y., Choudhari, S., Strahan, G. D., Pfeffer, P. E., Kiers, E. T., and Bücking, H. (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *PNAS*, *109(7)*, 2666-2671.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., and Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, *13(5)*, 538-539.

Fowler, J. H., and Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *PNAS*, *107(12)*, 5334–5338.

Frey, B. S., and Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, *40*, 402–435.

Frijda, N. (2007). *The Laws of Emotion.* Mahwah, New Jersey: Erlbaum.

Gabaix, X., and Laibson, D. (2017). Myopia and discounting. CEPR Discussion Paper 11914.

Geng, J. J., and Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience and Biobehavioral Reviews*, *37*, 2608-2620.

Gershman, S. J., and Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual Review of Psychology*, *68*, 101–28.

Gibbs, B. G., Forste, R., and Lybbert, E. (2018). Breastfeeding, parenting, and infant attachment behaviors. *Maternal and Child Health Journal*, *22(4)*, 579-588.

Gilboa, I., and Schmeidler, D. (1995). Case-based decision theory. *Quarterly Journal of Economics*, *110(3)*, 605-639.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *Quaterly Journal of Economics*, *115*, 811-846.

Glimcher, P. W. (2011). *Foundations of neuroeconomic analysis.* Oxford: Oxford University Press.

Glimcher, P. W., and Fehr E. (Eds.) (2014). *Neuroeconomics* (second edition). London: Academic Press.

Gneezy, A., and Fessler, D. M. T. (2012). Conflict, sticks and carrots: war increases prosocial punishments and rewards. *Proceedings of the Royal Society B*, *279*, 219–223.

Gordon, I., Vander Wyk, B. C., Bennett, R. H., Cordeaux, C., Lucas, M. V., Eilbott, J. A., Zagoory-Sharonc, O., Leckmand, J. F., Feldman, R., and Pelphrey, K. A. (2013). Oxytocin enhances brain function in children with autism. *PNAS*, *110(52)*, 20953–20958.

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*, 481–510.

Greiff, M., Ackermann, K. A., and Murphy, R. (2016). The influences of social context on the measurement of distributional preferences. MAGKS Joint Discussion Paper Series in Economics, No. 06-2016.

Grimmelikhuijzen, C. J. P., and Hauser, F. (2012). Mini-review: The evolution of neuropeptide signaling. *Regulatory Peptides*, *177*, S6–S9.

Gruber, C. W. (2014). Physiology of invertebrate oxytocin and vasopressin neuropeptides. *Experimental Physiology*, *99(1)*, 55–61.

Guzmán, Y. F., Tronson, N. C., Jovasevic, V., Sato, K., Guedea, A. L., Mizukami, H., Nishimori, K., and Radulovic, J. (2013). Fear-enhancing effects of septal oxytocin receptors. *Nature Neuroscience*, *16*, 1185-1187.

Halevy, N., Bornstein, G., and Sagiv, L. (2008). “in-group love’’ and ‘‘out-group hate’’ as motives for individual participation in intergroup conflict. *Psychological Science*, *19(4)*, 405-411.

Hauge, K. E., Brekke, K. A., Johansson, L., Johansson-Stenman, O., and Svedsäter, H. (2016). Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Economics*, *19*, 562–576.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., and Singer, T. (2010). Neural responses to ingroup and outgroup members’ suffering predict individual differences in costly helping. *Neuron*, *68*, 149–160.

Hershfield, H. E. (2011). Future self-continuity: how conceptions of the future self transform intertemporal choice. *Annals of the New York Academy of Sciences*, *1235*, 30–43.

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

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O’Doherty, J. P., and Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience*, *20*, 1142–1149.

Hoeve, M., Stams, G. J. J. M., van der Put, C. E., Semon Dubas, J., van der Laan, P. H., and Gerris, J. R. M. (2012). A meta-analysis of attachment to parents and delinquency. *Journal of Abnormal Child Psychology*, *40*, 771–785.

Houser, D., and Kurzban, R. (2002). Revisiting kindness and confusion in public good experiments. *American Economic Review*, *92(4)*, 1062-1069.

Insel, T. R., and Young, L. J. (2001). The neurobiology of attachment. *Nature Reviews Neuroscience*, *2*, 129-136.

Johnson, N. D., and Mislin, A. A. (2011). Trust games: A meta-analysis, *Journal of Economic Psychology*, *32*, 865–889.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Strauss, and Giroux.

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, *112(2)*, 375-404.

Kelly, A. M., and Vitousek, M. N. (2017). Dynamic modulation of sociality and aggression: an examination of plasticity within endocrine and neuroendocrine systems. *Philosophical Transactions of the Royal Society B*, *372*: 20160243.

Kemp, A. H., and Guastella, A. J. (2011). The role of oxytocin in human affect a novel hypothesis. *Current Directions in Psychological Science*, *20*, 222–231.

Kiers, E.T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., Fellbaum, C. R., Kowalchuk, G. A., Hart, M. M., Bago, A., Palmer, T.M., West, S. A., Vandenkoornhuyse, P., Jansa, J., and Bücking, H. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*, *333*(*6044*), 880-882.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435(2)*, 673-676.

Kramer J and Meunier J. (2016). Kin and multilevel selection in social evolution: a never-ending controversy? *F1000Research*, *5*, 1-13.

Kummel, M., and Salant, S. W. (2006). The economics of mutualisms: optimal utilization of mycorrhizal mutualistic partners by plants. *Ecology*, *87(4)*, 892-902.

LaBar, K. S., and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuro*, *7*, 54-64.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46(4)*, 352–367.

LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.

Leider, S., Möbius, M. M., and Do Rosenblat, Q.-A. (2009). Directed altruism and enforced reciprocity in social networks. *Quarterly Journal of Economics*, *124*, 1815–1851.

Lengyel, M., and Dayan, P. (2007). Hippocampal contributions to control: the third way. Conference: Advances in Neural Information Processing Systems 20, Vancouver.

Liakoni, V., Modirshanechi, A., Gerstner, W., and Brea, J. (2021). Learning in volatile environments with the Bayes factor surprise. *Neural Computation*, to appear.

Liberman, V., Samuels, S. M., and Ross, L. (2004). The name of the game: predictive power of reputations versus situational labels in determining prisoner’s dilemma game moves. *Personality and Social Psychology Bulletin*, *30(9)*, 1175-1185.

Liebrand, W. (1984). The effect of social motives, communication and group sizes on behaviour in an N-person multi-stage mixed motive game. *European Journal of Social Psychology*, *14*, 239–264.

Lijphart, A., 1975. *The Politics of Accommodation: Pluralism and Democracy in The Netherlands*. Second Edition, Revised. University of California Press, Berkeley.

Lind, S. E. (2010). Memory and the self in autism: a review and theoretical framework. *Autism*, *14(5)*, 430-456.

Lind, S. E., and Bowler, D. M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology*, *119(4)*, 896–905.

Loerakker, B., Bault, N., Hoyer, M., and van Winden, F. (2016). Asymmetry in the development of cooperative and antagonistic relationships: A model-based analysis of a fragile public good game experiment. CREED Working Paper, Amsterdam School of Economics, University of Amsterdam.

Loerakker, B., and van Winden, F. (2017). Emotional leadership in an intergroup conflict game experiment. *Journal of Economic Psychology*, *63*, 143–167.

Loewenstein, G. (1996). Out of control: visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65(3)*, 272–292.

Malmendier, U., and Schmidt, K. (2012). You owe me. CEPR Discussion Paper No. 9230.

Malmendier, U., and Schmidt, K. (2017). You owe me. *American Economic Review*, *107(2)*, *493–526.*

Massen, J. J. M., Sterck, E. H. M., and de Vos, H. (2010). Close social associations in animals and humans: functions and mechanisms of friendship. *Behaviour*, *147*, 1379-1412.

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503-507.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., and Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*, 73–79.

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio*, *3(2)*, 1-7.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *Quarterly Journal of Economics*, *117(3)*, 735-774.

Nelson, E. E., and Panksepp, J. (1998). Brain substrates of infant–mother attachment: contributions of opioids, oxytocin, and norepinephrine. *Neuroscience and Biobehavioral Reviews*, *22(3)*, 437–452.

Newman, L., Sivaratnam, C., and Komiti, A. (2015). Attachment and early brain development – neuroprotective interventions in infant–caregiver therapy. *Translational Developmental Psychiatry*, *3(1)*, 28647, 1-12.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314(1560)*, 1560-1563.

Numan, M. (2015). *Neurobiology of Social Behavior.* London: Academic Press.

Numan, M. (2016). Brain circuits for parental behavior and love, with implications for other social bonds. *Emotion Researcher*, January.

Numan, M., and Young, L. J. (2016). Neural mechanisms of mother-infant bonding and pair bonding: Similarities, differences, and broader implications. *Hormones and Behavior*, *77*, 98–112.

O'Connell, G., Christakou, A., Haffey, A. T., and Chakrabarti, B. (2013). The role of empathy in choosing rewards from another's perspective. *Frontiers in Human Neuroscience*, *7(174)*, 1-5.

Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., Bartz, J. A., Yee, J. R., and van Zuiden, M. (2013). The role of oxytocin in social bonding, stress regulation and mental health: An update on the moderating effects of context and interindividual differences. *Psychoneuroendocrinology*, *38*, 1883—1894.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: Cambridge University Press.

Ostrom, E., Gardner, R., and Walker, J. (1994). *Rules, Games, and Common-Pool Resources.* Ann Arbor: University of Michigan Press.

Ostrom, E. (2010). Beyond markets and states: polycentric governance of complex economic systems. *American Economic Review*, *100*, 641–672.

Ostrom, E. (2012). Coevolving relationships between political science and economics. *RMM*, *3*, 51–65.

Oswald, A. J. (1997). Happiness and economic performance. *Economic Journal*, *107(445)*, 1815–31.

Palombo, D. J., Keane, M. M., and Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. *Hippocampus*, *25*, 345–353.

Pan, X. S., and Xiao, E. (2016). It's not just the thought that counts: An experimental study on the hidden cost of giving. *Journal of Public Economics*, *138*, 22–31.

Parfit, D. (1971). Personal identity. *Philosophical Review*, *80(1)*, 3-27.

Pedersen, C. A. (2004). Biological aspects of social bonding and the roots of human violence. *Annals of the New York Academy of Sciences*, *1036*, 106–127.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M., and Bonhoeffer, S. (2005). *Proceedings: Biological Sciences*, *272(1568)*, 1115-1120.

Piff, P. K., Kraus, M. W., Côté, S., Hayden Cheng, B., and Keltner, D. (2010). The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99(5)*, 771–784.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., and Keltner, D. (2012). Higher social class predicts increased unethical behavior. *PNAS*, *109(11)*, 4086–4091.

Pitcher, D., Japee, S., Rauth, L., and Ungerleider, L. G. (2017). The superior temporal sulcus Is causally connected to the amygdala: a combined TBS-fMRI study. *The Journal of Neuroscience*, *37(5)*, 1156–1161.

Pronin, E., Olivola, C. Y., and Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: psychological distance and decision making. *Personality and Social Psychology Bulletin*, *34(2)*, 224-236.

Queller, D. C., and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B*, *364*, 3143–3155.

Quintana, D. S., and Guastella, A. J. (2020). An allostatic theory of oxytocin. *Trends in Cognitive Sciences*, *24(7)*, 515-528.

Rajan, R. (2019). *The Third Pillar*. London: William Collins.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677.

Rand, D. G., and Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS ONE*, *9*, e109687.

Reiss, H. T., Collins, W. A., and Berscheid, E. (2000). The relationship context of human behavior and development. *Psychological Bulletin*, *126*, 844–872.

Roth, A. E., and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, *8*(*1*), 164-212.

Scherer, K. R., Schorr, A., and Johnstone, T. (Eds.) (2001). *Appraisal Processes in Emotion.* Oxford: Oxford University Press.

Schino, G., and Aureli, F. (2009). Reciprocal altruism in primates: Partner choice, cognition, and emotions. *Advances in the Study of Behavior*, *39*, 45–69.

Schino, G., and Aureli, F. (2010). Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology*, 19, 130–135.

Schultz, W., Dayan, P., and Montague, R. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–99.

Seyfarth, R. M., and Cheney, D. L. (2012). The evolutionary origins of friendship. *Annual Review of Psychology*, *63*, 153–177.

Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*, 617–627.

Shohamy, D., and Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron, 60*, 378–389.

Silk, J.B. (2002). Using the 'f'-word in primatology. *Behaviour*, *139*, 421-446.

Simms, E. L., and Lee Taylor, D. (2002). Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative* Biology, *42*, 369–380.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, *30*, 855–863.

Soares, M. C., Bshary, R., Mendonça, R., Grutter, A. S., and Oliveira, R. F. (2012). Arginine vasotocin regulation of interspecific cooperative behaviour in a cleaner fish. *PLoS ONE*, *7(7)*, e39583, 1-10.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature*, *43*, 392–436.

Sonnemans, J., van Dijk, F., and van Winden, F. (2006). On the dynamics of social ties structures in groups. *Journal of Economic Psychology*, *27*, 187–204.

Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., and Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Science Advances*, *2(10)*, e1600992.

Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., and Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *PNAS*, *112 (5)*, 1619-1624.

Tankersley, D., Stowe, C. J., and Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nature Neuroscience*, *10*, 150–151.

Tennie, C., Jensen, K., and Call, J. (2016). The nature of prosociality in chimpanzees. *Nature Communications*, *7*, 13915, 1-8.

Teo, A. R., Choi, H., and Valenstein, M. (2013). Social relationships and depression: ten-year follow-up from a nationally representative study. *PLoS ONE 8(4)*, e62396, 1-8.

Tversky, A. (1977). Features of similarity. *Psychological Review, 64(4)*, 327-352.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211(4481)*, 453-458.

Uzzi, B. (1999). Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, *64*, 481–505.

van Dijk, F., Sonnemans, J., and van Winden, F. (2002). Social ties in a public good experiment. *Journal of Public Economics*, *85*, 275–299.

van Dijk, F., and van Winden, F. (1992). Dynamics of social ties and local public good provision. Faculty of Economics and Econometrics, University of Amsterdam, Research Memorandum 9226.

van Dijk, F., and van Winden, F. (1997). Dynamics of social ties and local public good provision. *Journal of Public Economics*, *64*, 323–341.

van Hooff, J. A. R. A. M. (2001). Conflict, reconciliation and negotiation in non-human primates: the value of long-term relationships. In R. Noë, J. A. R. A. M. van Hooff, and P. Hammerstein (Eds.), *Economics in Nature.* Cambridge: Cambridge University Press.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*, 337–349.

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., and Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*, 733–746.

van Veelen, M., Garcıá, J., and Avilés, L. (2010). It takes grouping and cooperation to get sociality. *Journal of Theoretical Biology*, *264*, 1240–1253.

van Winden, F. (2012). Affective social ties—missing link in governance theory. *RMM*, *3*, 108–122.

van Winden, F. (2015). Political economy with affect: on the role of emotions and relationships in political economics. *European Journal of Political Economy*, *40*, 298–311.

Vohs, K. D., Mead, N. L., and Goode, M. R. (2006). The psychological consequences of money. *Science*, *314(5802)*, 1154-1156.

Vrticka, P., and Vuilleumier, P. (2012). Neuroscience of human social interactions and adult attachment style. *Frontiers in Human Neuroscience*, *6*(*212*), 1-17.

Weisel, O., and Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*, *60*, 110–120.

West, S. A., and Kiers, E. T. (2009). Evolution: what is an organism? *Current Biology*, *19(23)*, R1080-R1082.

West, S. A., Kiers, E. T., Simms, E. L., and Denison, R. F. (2002). Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. Lond.* B, *269*, 685-694.

Whiteley, M., Diggle, S. P., and Greenberg, E. P. (2017). Bacterial quorum sensing: the progress of an emerging research area. *Nature*, *551*(*7680*), 313-320.

Wiegman, I. (2019). Payback without bookkeeping: The origins of revenge and retaliation. *Philosophical Psychology*, *32*, 1100-1128.

Yonelinas, A. P., and Ritchey, M. (2015). The slow forgetting of emotional episodic memories: an emotional binding account. *Trends in Cognitive Sciences*, *19(5)*, 259-267.

Young, L. J,, and Wang, Z. (2004). The neurobiology of pair bonding. *Nature Neuroscience*, *7(10)*, 1048-54.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, *39*, 117–123.

Appendix

Derivation of eqs. (2) – (4)

(For similar applications in other learning models in decision neuroscience and economics, see: Dayan et al. 2000, Behrens et al. 2007, Daw 2014, Gabaix and Laibson 2017.)

Let $\alpha'_{ijt+1} = \alpha_{ijt+1} - \alpha_{ijt}$ and $I'_{ijt} = I_{ijt} - \alpha_{ijt}$. Given I'_{ijt}, the distribution of α'_{ijt+1} is Gaussian and can be represented by

$$\text{(A1)} \quad \alpha'_{ijt+1} = \delta_{ijt} I'_{ijt} + \xi_t$$

for some δ_{ijt}, and some independent distributed noise term ξ_t, with variance σ^2_{ijt+1}. Multiplying both sides of (1) by I'_{ijt} and taking expectations, gives: $E\left[\alpha'_{ijt+1} I'_{ijt}\right] = \delta_{ijt} E\left[I'^2_{ijt}\right]$; thus,

$$\text{(A2)} \quad \delta_{ijt} = \frac{E\left[\alpha'_{ijt+1} I'_{ijt}\right]}{E\left[I'^2_{ijt}\right]} = \frac{E\left[\alpha'_{ijt+1}\left(\alpha'_{ijt+1} + \varepsilon_t\right)\right]}{E\left[\left(\alpha'_{ijt+1} + \varepsilon_t\right)^2\right]} = \frac{E\left[\alpha'^2_{ijt+1}\right]}{E\left[\alpha'^2_{ijt+1} + \varepsilon^2_t\right]} = \frac{\sigma^2_{ijt}}{\sigma^2_{ijt} + \sigma^2_\varepsilon}$$

Next, taking the variance of both sides of $(\mathrm{P}1)$, gives: $\sigma^2_{ijt} = \delta^2_{ijt}\sigma^2_{I'}{}_t + \sigma^2_{ijt+1}$, with: $\sigma^2_{I'}{}_t \equiv \sigma^2_{ijt} + \sigma^2_\varepsilon$.

Using (2), $\delta_{ijt}\sigma^2_{I'} = \sigma^2_{ijt}$, and, thus,

$$\text{(A3)} \quad \sigma^2_{ijt+1} = \sigma^2_{ijt} - \delta^2_{ijt}\sigma^2_{I'} = \sigma^2_{ijt} - \delta_{ijt}\sigma^2_{ijt} = \left(1 - \delta_{ijt}\right)\sigma^2_{ijt}$$

Hence, $\alpha_{ijt+1} \sim \mathcal{N}\left(\alpha_{ijt} + \delta_{ijt}\left(I_{ijt} - \alpha_{ijt}\right),\ \left(1 - \delta_{ijt}\right)\sigma^2_{ijt}\right)$, with $\delta_{ijt} = \sigma^2_{ijt} \,/\, \left(\sigma^2_{ijt} + \sigma^2_\varepsilon\right)$.