Forthcoming in the Journal of Economic Psychology
Special issue: Biological Foundations of Economic Decision Making

The Informational Affective Tie Mechanism

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

Frans van Winden
Emeritus Professor
Amsterdam School of Economics
and
Amsterdam Brain & Cognition
University of Amsterdam
f.a.a.m.vanwinden@uva.nl

ABSTRACT

Based on the growing evidence on caring and enduring relationships displayed by species across the evolutionary ladder, the ubiquity and importance of environmental uncertainty faced by all organisms, and the adaptational principle that learning may involve preference learning besides instrumental reinforcement learning, this paper proposes a novel information theoretic model of affective bonding, focusing on humans. A special case of the proposed "informational affective tie mechanism" (*i*ATM) turns out to be the model of Bault, Fahrenfort, Pelloux, Ridderinkhof, and van Winden: An affective social tie mechanism, Journal of Economic Psychology, 2017, 61, 152–175. In further contrast to the latter model, the *i*ATM model allows for the role of multiple contexts and distributed attention. Moreover, it provides a dynamic, context related, endogenous representation of the well-known social value orientation construct, facilitating the propagation of caring as observed in the literature. Empirical support is provided along different dimensions. Although the model is not estimated in full detail, a necessary condition regarding its parameters is shown to be fulfilled. Furthermore, experimental findings concerning various well-known games can be tracked under plausible calibration. In addition, the mechanism can be linked to neurobiological evidence concerning maternal (and paternal) care – as the presumed primordial caregiving system – and the signaling role of oxytocin. Finally, the evidence concerning non-human species is addressed, as well as the role of norms and reciprocity.

JEL classification: A13, C00, D01, D91, H41

Keywords: Affective ties; Uncertainty-based model; Social preference learning; Public good; Empirical support

1. Introduction

Affective social relationships or ties, that are emotional like friendships and involve care for another individual, abound and are widely deemed to be very important for an individual's welfare. Key characteristics of affective ties are that they are dynamic, based on emotional interaction experiences, and generalize across contexts (spread) and time (persistence). A friendship, for example, develops over time in terms of strength. The affective weight one attaches to a friend is not fixed from the start, but grows and may fluctuate. It may even grow negative, turning friendship into a hate relationship, with negative care instead of positive care for the other. Also, it is not restricted to the context it originates in (for instance, the workplace) but may extend to and further develop in other settings (like a sportsclub). Depending on the characteristics of a context (such as its hedonic value), the impact of an interaction experience on tie formation – its emotional imprint – is likely to differ. Think of a life-threatening context versus interaction at a gym, for instance. Moreover, one's (initial) attitude towards strangers in a new context may be influenced by previous interaction experiences, and the more so the greater the perceived similarity with the contexts of these experiences. Given our limited mental resources, finally, the differential attention that contexts attract in decision-making – which will be co-determined by their aforementioned differential imprint – is likely to impact choice behavior.

In a recent article in this journal Bault, Fahrenfort, Pelloux, Ridderinkhof, and van Winden (2017) refer to relevant literature on affective social relationships from across the social sciences, noting that existing formal models of social preferences miss out on one or more of the aforementioned characteristics. The main reason is that these models are typically static equilibrium models and, thus, do not address the aforementioned dynamics intrinsic to affective tie formation. Relatedly, there is no explicit consideration of multiple contexts and attention. The goal of this paper is to present a formal model that can account for all these characteristics or aspects, together with supportive evidence. The model proposed by Bault et al. (2017), to be discussed next, will be used as starting point.

Based on the theoretical model of van Dijk and van Winden (1997), Bault et al. (2017) propose and test an empirically implementable model that incorporates an affective tie mechanism (ATM, henceforth; see next section for a formal presentation). In their ATM model, an individual's care for another individual is formalized as a weight that the former attaches to the latter's welfare in decision-making, where they focus on voluntary contributions to a public good (benefiting all involved) as context. Crucially, this weight is not assumed to be fixed, but supposed to be determined by interaction experiences generating an affective tie (or bond) represented by the weight. Beneficial experiences feed the development of a positive tie, where the other (interaction counterpart) is perceived as a kind of friend. Harmful experiences, on the other hand, foster a negative tie, where the other is perceived as a foe. More specifically, the weight attached to another individual's welfare in decision-making – representing the current tie with that individual – is made up by a weighted combination of the tie that already existed (which may have a value of zero) and the interaction experience with that individual. The latter, called an impulse, is generated by the other's behavior in comparison with a reference behavior.

Thus, in addition to the two key parameters standing for the weights attached to the existing tie and the impulse (jointly determining the current affective tie), the ATM model comprises a reference behavior and an initial tie value. Whereas the reference behavior is taken to depend on the context, the well-known sociopsychological construct of social value orientation (SVO) is suggested as measure for the initial tie value. In support of this ATM model, Bault et al. (2017) provide empirical evidence from the estimation and predictive performance of the model regarding three different data sets of public good experiments (where behavior concerns the voluntary contribution to a

public good), a horse-race with other models (goodness-of-fit comparison), and a model-based fMRI analysis of brain activity data (taken from Bault, Pelloux, Fahrenfort, Ridderinkhof, & van Winden, 2015). Regarding the two key parameters in tie formation, they get an estimate of 0.5 for the weight related to an existing tie, and an estimate of 0.08 for the weight attached to an impulse. From the former estimate they conclude that "a decay of about 50% is observed".

Notwithstanding the positive results with the ATM model, the Bault et al. (2017) paper raises a number of important issues that need be addressed – and will be addressed in this paper – to strengthen the foundation and scope of the model. *First*, what determines the *key parameters* δ_{i1} and δ_{i2}? Does a tie value, furthermore, indeed decay over time? *Second*, what determines the *initial tie-value* α_{ij0}? And, why does the model-fit improve if SVO is taken as measure? *Third*, how to deal with *contexts*? SVO, for instance, appears to be context dependent (Bogaert, Boone, & Declerck, 2008; Murphy & Ackermann, 2014; Greiff, Ackermann, & Murphy, 2016). And, how to deal with *attention*, given multiple contexts and limited mental resources? *Fourth*, what is the relevance or applicability of the ATM for studying the behavior of *non-human species*? In passing, Bault et al. (2017) refer to studies on the evolutionary origins of affective bonding among *animals* and on mother-infant attachment as likely being the primordial caregiving mechanism that served as foundation for other types of prosocial bonding. They could have added recent studies regarding the biology of *plants*, *fungi*, and *bacteria* that are suggestive of similar processes occurring in these organisms (further discussed below). So, what unites these remarkably similar processes across the evolutionary ladder? Is there an underlying mechanism that we can model, with ATM as special case perhaps?

Note that this last issue involves species without higher-order cognitive skills that would enable calculated reciprocity, that is, deliberate strategic behavior (on the distinction between deliberation and affect, see: Kahneman, 2011; Loewenstein, O'Donoghue, & Bhatia, 2015). 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. But, what then is the driving factor that may further inform the nature of this mechanism? In this respect, the consideration that behavioral uncertainty regarding interaction counterparts is shared by all organisms strongly motivated the development of the uncertainty-related *informational* affective tie mechanism (*i*ATM) model proposed in this paper. It builds on the following hypothesis.

Basic Hypothesis 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.

Note in this context that, from an evolutionary perspective, there are two ways in which organisms can adapt to an uncertain environment: one way is to learn how to act on that environment (cf. standard instrumental reinforcement learning), while another adaptation – relevant here – concerns preference learning involving an internal state adjustment (Friston 2010). As in classical conditioning, both ways of adapting concern the learning of predictive relationships which deserve selective attention (Dayan, Kakade, & Montague, 2000). Of course, the precise nature of this hypothesized 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. Moreover, one should think of affect or emotion in an appropriate way, as further discussed below.

Based on these considerations, the *i*ATM model presented in this paper enables to address all the issues raised above regarding the ATM model. It offers a formal theoretical foundation for the

key parameters in tie formation as well as the initial tie value, allowing for multiple contexts attracting differential attention. More generally, it may provide a unifying formal framework for studying bonding across species.

The organization of this paper is further as follows. Section 2 presents and formalizes the informational affective tie mechanism (*i*ATM), including the role of contexts and attention. Section 3 goes into the empirical support for the model. Although it is beyond the scope of this paper to test the *i*ATM model in all detail, empirical support will be provided along several dimensions: (i) direct econometric evidence regarding a necessary condition with respect to the key parameters δ_{i1} and δ_{i2}; (ii) several applications where the model remarkably closely predicts important experimental game findings; (iii) supportive neurobiological evidence regarding parental care as exemplary case of bonding, including the role of oxytocin as signaling molecule; and, finally, (iv) SVO as practical measure of an initial tie-value. Section 4 discusses the relevance of the *i*ATM model for the study of other species, and shortly addresses two other topics of interest: norms and reciprocity. Section 5 closes with a concluding discussion.

2. The informational Affective Tie Mechanism

This section presents and formalizes the informational affective tie mechanism (*i*ATM). For convenience, a short summary of the Bault et al. (2017) ATM model follows first.

In the ATM model, an agent *i* may care for another agent *j*, that is, take *j*'s welfare or utility into account when choosing a behavior (action) from a behavioral repertoire (action set) that is available.[1] Bault et al. (2017) focus on the voluntary provision of a public good, in which case an action equals a contribution. Agent *i*'s care for *j*, at time *t,* is formalized as a weight α_{ijt} attached to *j*'s utility U_{jt}. This transforms *i*'s utility, U_{it}, into an extended utility $V_{it} = U_{it} + \alpha_{ijt}U_{jt}$. The value α_{ijt}, representing *i*'s current emotional or affective tie with *j*, determines together with *i*'s current interaction experience with *j*, where the latter is denoted as impulse I_{ijt} (further specified below), the updated tie value α_{ijt+1}. More specifically, the new tie is assumed to be a weighted combination of the existing tie and the current impulse: $\alpha_{ijt+1} = \delta_{i1}\alpha_{ijt} + \delta_{i2}I_{ijt}$, where the 'tie-persistence' parameter δ_{i1} and the 'tie-impulse' parameter δ_{i2} are (nonnegative) weights. An impulse, furthermore, is taken to equal the difference between *j*'s action a_{jt} and a reference action a_{jt}^{ref}, that is: $I_{ijt} = a_{jt} - a_{jt}^{ref}$. Thus, for example, in absence of an existing tie (thus, $\alpha_{ijt} = 0$) or a sufficiently small tie-persistence parameter δ_{i1}, a positive interaction experience ($I_{ijt} > 0, \delta_{i2} > 0$) will increase *i*'s tie with *j* , that is, $\alpha_{ijt+1} > \alpha_{ijt}$. A negative experience, on the other hand, may lead to a decrease in the tie value, which may even become negative. In case of a negative tie value *i* starts to care negatively about the welfare of *j*, and, instead of a willingness to help, a willingness to hurt develops (to be weighed against its cost). In addition to the two key parameters δ_{i1} and δ_{i2}, the ATM model comprises a reference action a_{jt}^{ref} and an initial tie value α_{ij0}.

2.1. The iATM Model

The *i*ATM model basically consists of 3 modules. Module 1 concerns a *friend-foe appraisal*, that is, an experiential assessment or evaluation of the true reward that is obtained contingent on interacting with a particular agent (the agent's type). A friend is associated with a predicted positive change in welfare or utility, while a negative change is associated with a foe. Or, put differently,

[1] Henceforth, 'agent' is used interchangeably with 'individual' or 'organism', and the same holds for 'welfare'and 'utility'.

friends (foes) are expected to care positively (negatively) about one's welfare. Module 2 deals with *affective tie formation*, formalizing the affective bond with the assessed type of the interaction counterpart, given the context attended to. Module 3, finally, regards a spillover or generalization effect, indicated as 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), given the interactions already experienced. It may be helpful to think again of the voluntary provision of a public good as the relevant context, with action referring to a contribution (other contexts are discussed below).

2.1.1. 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 for agent *i* contingent on 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 inverse of its value ($1/\sigma_\varepsilon^2$) can be seen as an indicator of *reliability*.

The model allows for the potential of efficiency, that is, the maximization of utility while internalizing behavioral externalities (the neglected impact of one's behavior on another agent's utility). Internalization happens if an agent attaches equal weight to an interaction counterpart's utility as to its own utility; in the running case, if $\alpha_{ijt} = 1$ and, thus, $V_{it} = U_{it} + U_{jt}$. To that purpose, the experienced impulse I_{ijt}, generated by *j*'s action a_{jt}, is normalized as follows:

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

where a_{ijt}^{ref} denotes a *reference action*, expected from a *j* who is neither friend nor foe (an uncaring type), and a_{ijt}^{eff} stands for an *efficient action*, that is, a cooperative action maximizing the joint welfare of *i* and *j* (in a context with externalities $a_{ijt}^{eff} \neq a_{ijt}^{ref}$). Notice 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) $\alpha_{ijt+1} = \alpha_{ijt} + \delta_{ijt}\left(I_{ijt} - \alpha_{ijt}\right) = \left(1 - \delta_{ijt}\right)\alpha_{ijt} + \delta_{ijt} I_{ijt}$,

and variance:

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

where:

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

Note from Eq. (2) that repeatedly cooperative behavior by *j* (that is, repeatedly, $I_{ij} = 1$) 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_{ij} – the *learning rate* – only depends on the ratio of behavioral uncertainty to type uncertainty ($\sigma_\varepsilon^2/\sigma_{ijt}^2$). This uncertainty 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)).

Neither the normalization (eq. (1)) nor the information extraction formalization (Eqs. (2)-(4)) is part of the ATM model of Bault et al. (2017), see above.

2.1.2. Module 2: Affective Tie Formation, Context, and Attention

The *key assumption* of this module – which distinguishes the model from more standard reinforcement learning – 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 (preference learning as adaptation through internal state adjustment).

The assessment of an agent's type may be more or less reliable due to type uncertainty (see above). 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: $\bar{\alpha}_{ijt} = f\left(\sigma_{ijt}^2\right)\alpha_{ijt}$, where $f\left(\sigma_{ijt}^2\right)$ represents *i*'s type risk preference or attitude as a function of type uncertainty σ_{ijt}^2. In case of *type-risk neutrality*, $f(\cdot)$ would be a constant function of type uncertainty, with the constant being equal to one: $f\left(\sigma_{ijt}^2\right) = 1$ (in which case $\bar{\alpha}_{ijt} = \alpha_{ijt}$). In contrast, *type-risk aversion*[2] would imply a negative first-order derivative, denoted by $f'(\cdot) < 0$ (and, thus, $\bar{\alpha}_{ijt} < \alpha_{ijt}$), while *type-risk seeking* would involve a positive first-order derivative, denoted by $f'(\cdot) > 0$ (and $\bar{\alpha}_{ijt} > \alpha_{ijt}$). For illustration, the following simple specification: $f\left(\sigma_{ijt}^2\right) = e^{-\sigma_{ijt}^2}$ could hold for risk-aversion, and $f\left(\sigma_{ijt}^2\right) = e^{\sigma_{ijt}^2}$ for risk-seeking. Note that, whatever the risk attitude, $\bar{\alpha}_{ijt} = \alpha_{ijt}$ if the appraisal of *j*'s type is assessed to be fully reliable ($\sigma_{ijt}^2 = 0$, and, thus, $f\left(\sigma_{ijt}^2\right) = e^0 = 1$). In case of type-risk neutrality, agents do not mind the risk, and behave *as if* their counterpart is fully reliable. Furthermore, under type-risk aversion, the affective 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* γ ($0 \leq \gamma \leq 1$) to an interaction context (more on this below).

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

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

Note that type-risk attitude and context-related attention are neglected in the model of Bault et al. (2017).

2.1.3. Module 3: Generalized Tie Value

[2] In a trust context, type-risk aversion may be related to betrayal aversion (Bohnet & Zeckhauser, 2004; Aimone, Ball, & King-Casas, 2015).

The third and final module addresses what happens if *i* subsequently meets an unrecognized agent *k* (a generalized other) in the same context. In that case, a natural assumption is that *i* will generalize its type appraisal based on the 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 this new agent, however, 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 (Eq. (5)) can be rewritten as:

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

while the GTV (Eq.(6)) 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 indistinguishable (i.e., perceived as identical, either because recognition is impossible or too demanding qua effort), other agents, whoever they are, would seem 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 action 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.

The GTV of the *i*ATM model endogenizes the unexplained initial tie value (α_{i0}) of the ATM model of Bault et al. (2017). The relationship with the static SVO measure as proxy for the initial tie value is discussed in the next section. Two other adaptations of the ATM model concern the next two topics: multiple contexts and uncertainty related volatility. At the end of this section it will be shown that the ATM model can be retrieved as a special case of the *i*ATM model.

2.2. Multiple 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, Wakker, & Sarin, 1997) of the interaction (determining its salience and emotional imprint). Among the relevant defining factors of a *context* are likely to be the following. First, the type of game that is played, with an important horizontal competition-cooperation dimension, and a vertical hierarchy or dominance dimension. Second, the type(s) of agent(s) involved, where uncertainty may be related to nature, culture, and existing ties with the protagonist. And, third, any other uncertainty influencing behavior apart from type uncertainty (like the behavioral uncertainty represented by σ_ε^2 in Module 1).

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 perceived similarity of the new context with earlier experienced contexts. Assuming that similarity, like timing and hedonic value for that matter, can be 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 (Eq. 5a) can be rewritten as:

(5b) $V_{it} = U_{it} + \gamma_{i\mathbb{C}1t} \sum_{c1\epsilon C1} \bar{\alpha}_{ic1t} U_{c1t} + \gamma_{i\mathbb{C}2t} \sum_{c2\epsilon C2} \bar{\alpha}_{ic2t} U_{c2t} + \cdots + \gamma_{i\mathbb{C}\mathrm{N}t} \sum_{c\mathrm{N}\epsilon C\mathrm{N}} \bar{\alpha}_{ic\mathrm{N}t} U_{c\mathrm{N}t}$,

with: $0 \leq \sum_{\mathbb{C}\epsilon\mathfrak{C}} \gamma_{i\mathbb{C}t} \leq 1$, while the generalized tie value (eq. 6a) becomes:

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

2.3. Uncertainty related volatility

Perceived uncertainty may be affected due to changes in the environment, for instance, the set of agents that can be met. An important research topic in this regard is the treatment of surprises, that is, unexpectedness in contrast to unlikeliness (see, e.g.: Faraji, Preuschoff, & Gerstner, 2018; Liakoni, Modirshanechi, Gerstner, & Brea, 2021). The latter is readily captured by probability distributions, as commonly assumed for uncertainty, turning uncertainty into risk (Knight, 1921). In Module 1 this practice was followed regarding type uncertainty and behavioral uncertainty. To allow for *volatility*, more specifically, the possibility of repeated random shocks to the true reward that can be expected from a counterpart, let: $\tau_{ijt} = \tau_{ijt-1} + \eta_t$, 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) $\sigma_{ijt+1}^2 = \big(1 - \delta_{ijt}\big)\big(\sigma_{ijt}^2 + \sigma_{shock}^2\big)$,

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

Note that both the variance and the learning rate are increased, which leads to greater reliance on current impulses relative to the existing tie value in case of a volatile environment (see Eq. (2)). Furthermore, $f\big(\sigma_0^2 + \sigma_{shock}^2\big)$ has to be substituted for $f(\sigma_0^2)$ in the GTV expression (Eq. (6b)):

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

which gets smaller (larger) for type-risk averse (type-risk seeking) agents.

2.4. ATM model as special case

The ATM model of Bault et al. (2017) can be retrieved as a special case of this paper's *i*ATM model. To get to the former requires: (i) an unnormalized impulse (eq. (1)); (ii) interaction in a dyad in a fully attended ($\gamma = 1$) single context (no spill-overs); (iii) type uncertainty and behavioral uncertainty being such that the weights (δ) given to the existing (prior) tie value and the current impulse stay the same; and, (iv) agents being type-risk neutral ($f(\cdot) = 1$).

3. Empirical support

Although it is beyond the scope of this paper to estimate the *i*ATM model in full, empirical support will be provided by, among other things, testing a necessary condition for this model to hold, predicting experimental findings of several important games (regarding public goods, appropriation, bargaining, and trust), and relating the model to novel neurobiological evidence regarding maternal care as exemplary case of affective tie formation (bonding), including the role of oxytocin as type-information transmitter.

3.1. Direct econometric evidence

A necessary condition for the *i*ATM model to hold is that the weights in the affective tie mechanism, attached to the existing tie value and an impulse add up to 1 (see eq. (2)). For the dataset of Bault et al. (2017) this would mean that the individual econometric estimates of δ_{i1} and δ_{i2} should add up to 1, once δ_{i2} gets multiplied by 7 to deal with the required normalization of impulses (Eq. (1)).[3] Note that these parameters were unrestricted in estimation (except for being nonnegative). Using the SVO measure[4] for the initial tie value, statistical testing shows that the median of the sum ($\delta_{i1} + 7 \cdot \delta_{i2}$) equals 1.02, while the hypothesis of this sum being equal to 1 cannot be rejected (t-test: p = 0.20; Wilcoxon signed-rank test: p = 0.13); see Fig.1.[5]

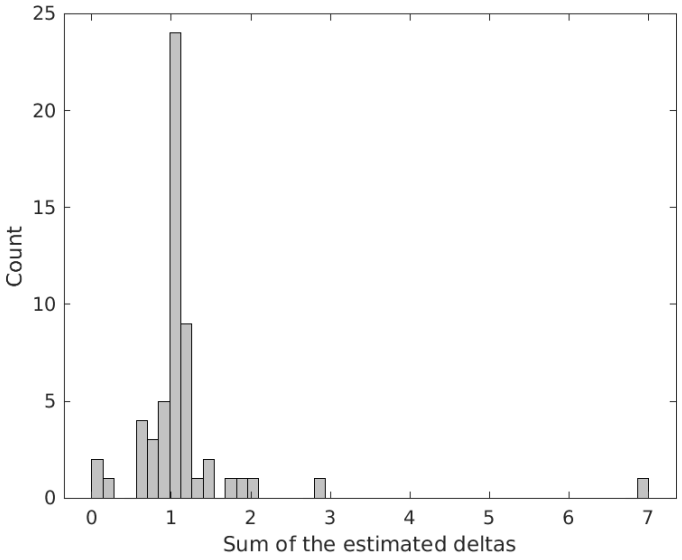

Fig. 1. Absolute frequency of the summed estimated deltas (after impulse normalization)

[3] The estimates of Bault et al. (2017) are based on a contribution of 3 as reference point (the standard Nash-equilibrium prediction, for which empirical support is provided), while the efficient contribution is 10; consequently, $a_{ijt}^{eff} - a_{ijt}^{ref} = 7$. As they use $I_{ijt} = a_{jt} - a_{jt}^{ref}$ (see Section 2) instead of $I_{ijt} = \left(a_{jt} - a_{ijt}^{ref}\right)/\left(a_{ijt}^{eff} - a_{ijt}^{ref}\right)$ their estimate of δ_{i2} gets divided by 7.

[4] Bault et al. (2017) use the Ring-test as continuous SVO measure. In a Ring-test, which measures distributional preferences, multiple choices between two alternatives, each representing a payoff allocation to Self and Other, have to be made. Moreover, each payoff combination (alternative) is taken from a circle. See Liebrand (1984). Each preferred payoff allocation can be considered as a vector. Added across choices, the angle of the resulting vector is an individual measure of the care for Other.

[5] Using 0 as initial tie value, instead of the SVO, renders similar results: a median of 1.05, p = 0.26 for the t-test and p = 0.15 for the Wicoxon signed-rank test. I am grateful to Nadège Bault (University of Plymouth) for providing me with these test results and the Figure.

This support for the *i*ATM model would imply that, in contrast to what is proposed in Bault et al. (2017), these weights are not stable personality traits, with δ_{i1} (labeled: tie-persistence parameter) assumed to be revealing the speed with which the tie decays 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. An implication that seems quite realistic (see, for instance, Delgado, Frank, & Phelps, 2005). A tie may not decay, furthermore, but leave a state in memory that remains eligibile for updating (an 'eligibility trace', see the text box by Niv in Glimcher & Fehr, 2014, pp. 305-306; see also Frijda, 1988, p. 354, on the power of emotional events to elicit emotions indefinitely). What would change instead is that memory retrieval becomes more effortful (and, thus, more costly) the longer a tie has not been activated through new impulses. The experience that nothing seems changed emotionally at a reunion with an old friend may count as anecdotal evidence.

3.2. Tracking experimental game findings

Because the iATM concerns an emotional behavioral motivation, the focus here will be on tracking responses generated in (repeated) one-shot games concerning voluntary public good provision, appropriation, bargaining, and investment. However, the *i*ATM model can be incorporated in a more general model allowing for strategic, forward-looking behavior (see: Bault et al., 2017; Loerakker, Bault, Hoyer, & van Winden, 2022), which would better fit the investigation of first-mover behavior. In all cases, type-risk neutrality and a fully attended context is assumed.

3.2.1. Public good game (propagation of caring)

In contrast to SVO, considered to reflect a (relatively) stable trait, the GTV construct of the *i*ATM model (see Eqs. (6)-(6b)) can foster a *propagation of caring* over time (persistence) as well as across individuals (spreading), induced by interaction experiences. Notably, the model fairly accurately predicts the "cooperative behavior cascades" findings of Fowler and Christakis (2010). Using data from Fehr and Gächter (2002) regarding repeated one-shot (linear) public good game experiments with groups of four (randomly and anonymously matched) 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 that the latter finding is only significant at the 20% level). To arrive at related *i*ATM model predictions the following assumptions are made: $\sigma_{ijt}^2 = \sigma_\varepsilon^2$ (applying the principle of insufficient reason, given different experiences in the repeated games), a reference contribution of 0 (standard Nash-equilibrium contribution), and a smooth utility function such that the contribution relative to the efficient contribution is determined by the tie value. Then, applying Eq. (6a), 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): $\left(\delta_{ijt} \,/\, 3\right)^{-t}$, with $\delta_{ijt} = ½$ 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 results are quite similar to the empirical observations of Fowler and Christakis (2010). 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, … : $\left(1-\delta_{ijt}\right)^{t-1}\left(\delta_{ijt}\,/\,3\right)$, with again $\delta_{ijt}=$ ½ 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 may take place.

3.2.2. Power-to-take game (appropriation)

In the power-to-take game (Bosman & van Winden, 2002) each of two randomly and anonymously matched players first earns a certain endowment *Z*. Subsequently, they are informed that one of them (the take authority, *i*) gets the opportunity to claim a share of the other's endowment, determining the take rate *t*: $0 \leq t \leq 1$). Finally, the latter (the responder, *j*) has the option to destroy any part of his or her *own* endowment, determining the destruction rate *d*: $0 \leq d \leq 1$). Although this is a very simple game, it captures some fundamental aspects of appropriation (like taxation). Under the standard (rational, selfish) homo economicus assumption of economic theory no destruction (*d* = 0) is predicted for the responder, while taking all (*t* = 1) is predicted for the take authority (or virtually all, leaving the smallest possible reward for the responder). In contrast, the experimental results show that on average *t* = 0.6 and *d* = 0.2, while responders expected a take rate of about 0.66 (t^{exp} = 0.66). Self-reported emotions of responders reveal that (high arousal) anger-related emotions play an important role in the decision to destroy, which typically entails destroying nothing (*d* = 0) or everything (*d* = 1).[6] This step-wise behavior is in line with psychological evidence showing that at higher intensities emotional urges progressively take control over behavior, rather than being compromised with what seems best according to a higher-order cognitive analysis of the consequences (here, losing everything). Further analysis of the probability of destruction (using a logit function; Bosman, Sutter, & van Winden, 2005, p. 420) indicates that responders become indifferent between destroying and keeping their after-the-take share (1−*t*)*Z* at *t* = 0.8.

The *i*ATM model can accurately generate this crucial take rate of *t* = 0.8. Using 0 for the efficient take rate (t^{eff} = 0, assuring the most money for both) and the expected take rate as reference ($t^{ref} = t^{exp}$ = 0.66) , the responder faces a normalized impulse equal to: $I_{ji} = (0.66 - t)$ / 0.66. For this one-shot (unrepeated) game, it seems plausible to assume that type uncertainty (σ_{ji}^2) regarding the take authority will be larger than behavioral uncertainty given its type (σ_ε^2), such that the learning rate (δ_{ji}) is close to 1. Then, with utility linear in payoff (in view of the high arousal step-wise behavior), the responder's extended utility in case of no destruction becomes: $V_j = (1-t)Z + \alpha_{ji}(1+t)$Z, with $\alpha_{ji} = (0.66 - t)\,/\,0.66$, while utility equals $\alpha_{ji}Z$ in case of destruction. Indifference, requiring equality of the two cases, is reached at a take rate of 0.82, or rounded: *t* = 0.8, as observed.

3.2.3. Ultimatum game (bargaining)

The power-to-take game is related to, but different from the well-known ultimatum bargaining game (Güth, Schmittberger, & Schwarze, 1982). In the latter one-shot two-player game, one of the two players (the proposer, *i*) gets an endowment, say Z, and can make a proposal how to share it with the other player (the responder, *j*). Subsequently, the only option for the responder is either to accept the proposal or to reject, in which case both get nothing. Experimental results show that responders, on average, expect a share of 50% (Chang & Sanfey 2009) and reject offers of

[6] Mediation analysis shows that the impact of the take rate on destruction is fully mediated by emotions (Bosman, Hennig-Schmidt, & van Winden, 2017).

around 20% with 50% chance (implying indifference; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003).

Because of the relatedness of this game with the power-to-take game, similar assumptions are used to apply the *i*ATM model, except that now the expected share claimed is 0.5 instead of 0.66. Thus, letting *s* denote the proposed share: $s^{ref} = s^{exp} = 0.5$, while $s^{eff} = 0$. Then, the responder faces a normalized impulse equal to: $I_{ji} = (0.5 - t)$ / 0.5. With $\delta_{ji} = 1$ this implies as tie value: $\alpha_{ji} = (0.5 - t)$ / 0.5, and as extended utility in case of no rejection: $V_j = (1 - s)Z + \alpha_{ji}(sZ)$, while utility becomes 0 in case of rejection. Consequently, indifference is here reached at a proposed share *s* of 29%, which is fairly close to the finding of 20%.

3.2.4. Trust game (investment)

In a one-shot two-player trust game (Berg, Dickhaut, & McCabe, 1995) both players get an equal endowment (*Z*, say). Then, one of them (the sender or trustor *i*) gets the opportunity to send an amount (say, *T*) out of her or his endowment ($0 \leq T \leq Z$) to the other player (the responder or trustee *j*). The amount sent is tripled by the experimenter (like a return on an investment) and the responder can return any part (*R*, say) of this tripled amount ($0 \leq R \leq 3T$). Whereas homo economicus is predicted to return nothing and, thus, no transfers would take place, experimental results show that on average transfers equal about 50% of the endowment: $T = 0.5Z$, while back-transfers are more or less equal to the transfer: $R = T$ (Berg et al., 1995; Glaeser, Laibson, Scheinkman, & Soutter, 2000; Camerer, 2003). Note that the efficient transfer equals the whole endowment ($T^{eff} = Z$).

For the application of the *i*ATM model similar assumptions are used as before, except that now little money is assumed to be expected by the responder: $T^{exp} = 0$. Consequently: $\alpha_{ji} = \mathrm{T} / \mathrm{Z}$. Another difference relates to utility. As receiving a transfer is less (negatively) emotional than the anger-arousing context of a power-to take game or ultimatum game, more room for deliberation by the responder may be expected (Fredrickson & Branigan, 2005). To allow for a more flexible (smoother) trade-off between the trustee's utility and trustor's utility, a simple loglinear instead of linear extended utility function is assumed: $V_j = ln(2.5Z - R) + \alpha_{ji} ln(0.5Z + R)$. Maximization of V_j leads to $R = T = 0.5Z$, in line with what is observed.[7]

3.3. Suggestive neurobiological evidence

The affective tie mechanism, as a mechanism of affective bonding, raises the question how it relates to recent neurobiological findings regarding maternal care (attachment) and the bonding of mammals (Insel & Young, 2001; Numan, 2015, 2016; Numan & Young, 2016; Feldman, 2016, 2017). Before going into this question, first, the evidence-based neural network model of maternal care proposed by Numan (2015, 2016) is summarized (see also Numan & Young, 2016). This maternal care system is suggested to provide the neural foundation of human bonding more generally, like friendships; (Numan, 2015, p. 271; Numan, 2020, p. 17). Then, it will be indicated how existing neural evidence related to the ATM model (Bault et al., 2015) fits into this neural network model. Finally, additional evidence supportive of the more general uncertainty related *i*ATM model will be put forward.

3.3.1. Numan's (2015, 2016) neural network model of maternal care

[7] A meta-analysis by Johnson and Mislin (2011), covering 162 replications of the trust game from all over the world, finds again an overall-average transfer of 0.5, but a lower overall-average return as percentage of the responder's total wealth ($Z + 3T = 2.5Z$) of 0.372. The equivalent *i*ATM model prediction ($R/2.5Z$) equals 0.2.

Starting point in Numan's (2015, 2016) 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. 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 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. It will be returned to 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 love and empathy, this hypothetical neural model (see also Numan, 2020) is extended with some additional brain areas. Love and emotional empathy (sharing other's feelings) are associated with an area involving the anterior insula (AI), while cognitive empathy (understanding other's feelings) as well as mentalizing (understanding other's mental state, for example, thoughts or intentions) are associated with the temporoparietal junction (TPJ) and the neigboring superior temporal sulcus (STS). 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. Along this circuitry, interaction between cognitive empathy and emotional empathy may lead to empathic care (Ashar, Andrews-Hanna, Dimidjian, & Wagner, 2017), that is, a motivation to care for the other's welfare in a way that is felt and deemed appropriate (Numan, 2020, p. 248).

Because males are not similarly exposed to pregnancy hormones, paternal care cannot be explained along the same lines. In this case, Numan (2020) 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.

3.3.2. Mapping the affective tie mechanism onto the neural network model

Bault et al. (2015) show the results of a model-based brain imaging (fMRI) study of the ATM model, using the same parameter estimates as in Bault et al. (2017). Their main findings are the following. First, at the feedback phase in a round of the public good game, when participants learn their partner's contribution, brain activity in the STS (plus the neighboring TPJ) and AI is related to the impulse I_{ijt}. Note that the former is not only implicated in mentalizing and cognitive empathy, but also in inferring the relevance of others and the signaling of cooperative partners and friends[8], while the latter is more generally implicated in emotions (see Bault et al., 2015). Second, at the decision phase, when the contribution decision is made: (1) the tie value α_{ijt} appears related to activity of the STS; (2) the parameter estimates of δ_{i1} and δ_{i2} are related to activity in the same region (STS); (3) the STS, furthermore, appears to be functionally connected with the mPFC; while (4) activity in the mPFC and AI is, in turn, modulated by the contribution magnitude.

These findings fit and add to Numan's (2020) neural network model of parental caring. They fit, in particular, because of the observed role of interaction experience and the connectivity between the STS, the mPFC, and the amygdala (see also: Decety & Svetlova, 2012; Bickart, Dickerson, & Feldman Barrett, 2014; Pitcher, Japee, Rauth, & Ungerleider, 2017), with potential modulation of the STS by OT (Bethlehem, van Honk, Auyeung, & Baron-Cohen, 2013; Gordon et al., 2013; Abraham et al., 2014). Importantly, these findings also add the affective tie mechanism (the tie value) as likely source of empathy, which is left unexplained in Numan's (2020) model. As empathic concern or care requires effort, and thus mental resources, it seems not automatic but dependent on the valuation of the other (Singer, 2006; Batson, Eklund, Chermok, Hoyt, & Ortiz, 2007; Hein, Silani, Preuschoff, Batson, & Singer, 2010; Decety & Svetlova, 2012; see also Fahrenfort, van Winden, Pelloux, Stallen, & Ridderinkhof, 2012). Therefore, the following adaptation of the Numan (2020) 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 STS as integrator of information 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 (AI, implicated in empathy) showed a positive parametric modulation by the contribution magnitude during the decision phase in the experiment of Bault et al. (2015) discussed above. Finally, to mention just one consequence of adding the tie mechanism, note that even if an infant stimulus (impulse) would be perceived as negative it need not turn a caretaker's positive care into negative care (see Eq. (2)), as the updated tie value may still be positive.

3.3.3. The role of OT in type prediction and type prediction error

The interaction between OT and DA along the circuitry is considered to be critical in the Numan (2015, 2020) model. Recent studies suggesting the involvement of OT in uncertainty reduction may provide an even deeper link between this model and the *i*ATM model. Starting point are the following four observations in the recent literature. *Firstly*, OT is not only involved in facilitating prosocial behavior, as often suggested, but also in facilitating antisocial behavior (De Dreu, 2012; Olff et al., 2013; Guzmán et al., 2013; Kemp & Guastella, 2011; Kelly & Vitousek, 2017).

[8] Geng and Vossel (2013) propose "contextual updating" as general characteristic of this area: "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" (Geng & Vossel, 2013, p. 2617).

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, Monakhov, Maayan, & Ebstein, 2016; Quintana & 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 (2015, 2020) neural network model. Whereas DA is related to reward prediction and reward prediction error, irrespective of the stimulus type (Schultz, Dayan, & Montague, 1997; Montague, Dayan, & Sejnowski, 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 & Schultz, 2013).

To illustrate how the *i*ATM model might fit into this picture, assume again a two-person (*i* and *j*) public good game context. Let the game for protagonist *i* start 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 ($I_{ijt} - \alpha_{ijt}$), 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 STS (and neighboring TPJ), where assumedly the activity of a population of neurons reflecting the type distribution (and the uncertainty related learning rate δ_{ijt}) is adjusted via neural plasticity, generating a new tie value α_{ijt+1}. 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; see Brown & 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 AI, 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_{ijt} as prior from the STS.

3.4. GTV as uncertainty-based formal underpinning of SVO

SVO is usually measured by having people decide between a small number of alternatives regarding payoff allocations to the self and another person – typically related to a social dilemma context –, whereafter respondents are classified as cooperators, individualists or competitors (e.g.: Van Lange, Otten, De Bruin, & Joireman, 1997; Bogaert et al., 2008; Murphy & Ackermann, 2014). SVO shows a significant small to medium effect size for cooperation in social dilemmas (Balliet, Parks, & Joireman, 2009).

Problematic from a theoretical and empirical point of view, however, is the lack of a unifying overarching theory, that findings are incompatible with a categorical conceptualization of SVO, and that the static stable trait approach is inadequate in accounting for how individual preferences

change in different situations and contexts (Murphy & Ackermann, 2014, p. 15). For instance, because of the categorization, (within or across category) adaptation of individual preferences is typically missed. In this respect, another SVO measure, using the Ring-test referred to above (see footnote 5), performs better because its angle measure provides a continuous measure of types (see also Murphy, Ackermann, & Handgraaf, 2011).

Theoretically, on all these scores, the dynamic, continuous, context related, and information theory based GTV construct of the *i*ATM model seems more satisfactory. Evidence supporting this construct relates to the following. Changes in the individual SVO (angle) are observed after social interaction experiences in a social dilemma experiment (Brandts, Riedl, & van Winden, 2009) and in a public good experiment (Ackermann & Murphy, 2019), while van Dijk, Sonnemans and van Winden (2002) find no impact of an individual decision-making task. In contrast to the static SVO, the dynamic GTV construct helps explain the "propagation of caring" (its persistence and spread) discussed above. It clarifies, moreover, why linking adult SVO to parental care related attachment styles (Van Lange et al., 1997) need not be successful (Ijzerman & Denissen, 2019). Although childhood experiences may blaze the trail, they can be obliterated by interaction experiences later in life. The role played by contexts, furthermore, finds support in the observed context dependency in the measurement of SVO (Greiff et al., 2016; Bogaert et al., 2008; see also below on reciprocity). Finally, the influence of information extraction is suggested by the observation in van Dijk et al. (2002) of a, relative to the outcomes in the final rounds, stronger impact of the SVO in the regression of the post-interaction angle regarding the interaction partner (representing the tie value for that partner). According to the *i*ATM model, this may be due to less type-uncertainty in the later rounds of the experiment, with a relatively smaller impact of the impulses in these rounds as consequence (see Eqs. (2) and (4)), while SVO picked up the impact of the GTV in the early rounds.

Nevertheless, as practical measure of the GTV an individual continuous SVO measure like the Ring-test angle may be useful if the task is applied before a related social dilemma decision-making context that participants are made aware of (so that the GTV that is subsequently elicited by the decision-making context is already elicited at the SVO task). In principle, the measurement should be repeated before interaction with new others in the same context to account for the impact of interaction experiences.

4. Other species, norms and reciprocity

4.1. Other species

Growing evidence on prosocial behavior and enduring relationships among very different animal species, and even plants, fungi, and bacteria, suggests the potential relevance and applicability of the *i*ATM model to a much wider range of organisms/agents. In fact, this is to be expected given the ubiquitous challenge faced by organisms of adapting to the behavioral uncertainty in interactions with other organisms.

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 (if friend) or negative (if foe) 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 (Cronin, 2012).

4.1.1. Enduring relationships among animals, plants, fungi, and bacteria

There are many examples of animals showing prosocial behavior – improving another individual's welfare – and enduring relationships (partnerships, bonds, sometimes called friendships; see: Massen, Sterck, & Vos, 2010; Seyfarth & Cheney, 2012). Among them are primates like chimpanzees and baboons, horses, dolphins, elephants, hyenas (Seyfarth & Cheney, 2012), cows (de Freslon, Peralta, Strappini, & Monti, 2020), rodents such as voles (Young & Wang, 2004) and rats (Ben-Ami Bartal, Rodgers, Sol Bernardez Sarria, Decety, & Masson, 2014), birds like parrots (Brucks & von Bayern, 2020), and fish (Soares, Bshary, Mendonça, Grutter, & Oliveira, 2012). Evidence of prosocial behavior and enduring relationships extends to plants, fungi, and bacteria, as indicated by studies on mutualisms between plants and mycorrhizal fungi (Kummel & Salant, 2006; Kiers et al. 2011; Fellbaum et al., 2012) and between legumes and rhizobia (Simms & Taylor, 2002; West, Kiers, Simms, & Denison, 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 (as they do with 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 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 and the uncertainty related underpinning. 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 (Schino & Aureli, 2009, p. 53). 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 (Schino & Aureli, 2009, p. 54). 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 (Schino & Aureli, 2009, p. 55). 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." (Schino & Aureli, 2009, p. 59). They note that emotional mediation makes long-term reciprocity possible (Brosnan & 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 & Schaffner, 2002; Schino & Aureli, 2010; Evers, de Vries, Spruijt, & Sterck, 2015, 2016), and shows clear parallels with the affective tie formation part of the *i*ATM model.

Of course, in relation to plants, fungi, and bacteria one should think of emotions and affect in an appropriate way. Appraisal theory of emotion (Lazarus, 1991; Scherer, Schorr, Johnstone, 2001; Frijda, 2007) offers a theoretical window allowing for translational continuity (de Waal, 2008) by viewing emotions as being determined by the evaluation (appraisal) of an event or behavioral episode, which can be more or less refined, and does not need to involve consciousness (Aureli & Schaffner, 2002). This way it seems possible 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 construct 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. OT-like peptide signaling systems appear to be more than 600 million years old (Grimmelikhuijzen & Hauser, 2012; Gruber, 2014; Feldman et al., 2016; Quintana & Guastella, 2020). These peptides 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 reproductive behavior (such as mate recognition), learning and memory (Gruber, 2014). It is expected, therefore, 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 OT 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 (recall the discussion of the role of OT in Section 3). 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, Volkmann, Hlavacka, Mancuso, & Barlow, 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 behavior, and perhaps even to the behavior of plants, fungi, and bacteriae.[9] The advantages of having such a formal model are, among others:

[9] If true, this would seem to question 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, p. 271, including references). If the above argumentation regarding the evolutionary origin of the affective tie caring mechanism is correct, mother-infant bonding may, in fact, have piggybacked on this more fundamental mechanism, assisted by pregnancy hormones and opioids to provide a jumpstart for attachment with the fetal allograft (see: Nelson & Panksepp, 1998; Douglas & Russell, 2001).

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 just one example, Tennie, Jensen, and Call (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.

4.1.2. iATM 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 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 & 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 (Loerakker et al., 2022). Apart from reciprocity, *i*ATM seems consistent with various other rules for the evolution of cooperation distinguished in the literature (Nowak, 2006; Bowles & Gintis, 2011; Kramer & 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 (that is, $r = \alpha$). 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 neither 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 separate themes (van Veelen, Garcıá, & Avilés, 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 may generate cooperative actions and affective tie formation in the initial absence of affective ties. *First*, choices may be stochastic, as often assumed in empirical decision models. Consequently, cooperative actions may happen, which may trigger tie formation leading to mutual cooperation (as in a standard binary choice prisoner's dilemma game, see next Section). *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, Lenski, and Zinser (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. (2012)). *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 (Pedersen, 2004; see further below).

4.2. On norms and reciprocity

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 the behavior of an interaction counterpart. 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 deliberate (forward-looking, strategic) reciprocity. The decision-making impact of the *i*ATM may be influenced by such higher-order cognitive processes, for example, through self-control (the regulation of emotional urges). The supportive evidence reported in this paper, and for the restricted ATM version in Bault et al. (2017), shows that the *i*ATM model in itself has substantial bite already in explaining and predicting human behavior. Nevertheless, the decision-making influence of norms and deliberate reciprocity may be significant dependent on the availability of mental resources (think of cognitive load effects) and context. Regarding the latter, in sequential decision-making, for example, they may be more relevant for first movers, who are stimulated to look forward, than for responders for whom emotions (triggered by first-mover behavior) are likely to play a stronger role. Even so, experimental evidence suggests that planning ahead and acting strategically is severely limited (for a review, see Bault et al., 2017). A few further comments regarding norms and reciprocity are in order.

To start with norms, first notice that adherence to behavioral standards that is not intrinsically motivated but strategic cannot explain the costly prosocial or antisocial responder behavior observed in one-shot game experiments (see Section 3). Intrinsic motivation, on the other hand, requires in this case that internalization of a norm has taken place through the reward and punishment by caretakers or educators (who may be peers). Internalization implies that the adherence will occur even if not monitored, due to the anticipated social emotions like shame and pride. 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 may lead to the (anticipated) experience of shame or pride when the norm is violated or adhered to, respectively. If reward and punishment is applied by an aspirant norm sender that one does not positively care for, however, the adherence will only be strategic and externally motivated by the expected reward or punishment. In that sense, the instilling and internalization of social norms runs on the more basic software of the affective tie mechanism (see Burnham, McCabe, & Smith, 2000), and is therefore abstracted from in this paper.

Incidentally, evoking norms for explanation without further substantiation is problematic, for can be regarded as "deus ex machina". For human behavior norms may function as reference point in the stimulus of the *i*ATM model in appropriate contexts. Interestingly, furthermore, the intrinsic motivation to comply with a social norm is related to the connectedness with one's future self (due to the role of anticipated social emotions). This is supported by recent (neuro)psychological studies,

distinguishing between temporal selves, showing that the future self is perceived as another person with whom similar affective relationships – psychological closeness or connectedness – can be developed as with other people. For example, experimental evidence shows that decisions for future self, like savings, are similar to decisions for other people (Pronin, Olivola, & Kennedy, 2008; Bartels & Rips, 2010). Furthermore, brain activity triggered by thinking about the future self shows a similar pattern of activation as thinking about another person (Hershfield, 2011; Soutschek, Ruff, Strombach, Kalenscher, & Tobler, 2016). Soutschek et al. (2016) find that the same brain region that plays a prominent role in the affective tie mechanism, the 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). See the more extensive working paper van Winden (2021), for further discussion on the relationship between social preferences, time preferences, as well as risk preferences, linked through uncertainty.

Regarding reciprocity, note that even though the affective tie mechanism bears some relationship with the common use of this term in behavioral economics for describing motivations to reward kind actions and punish unkind ones (e.g., Falk & Fischbacher, 2006), it differs in four important respects. *First*, although a kind action – generating a positive impulse – is a positive input for the tie value (see Eq. (2)), and the reverse holds for an unkind action, it need not lead to, respectively, reward and punishment. The reason is that the tie value is a stock variable, which means, for instance, that an unkind action may still leave a positive tie value (as in friendship), with no reciprocation as consequence. It is only in the absence of an existing tie and if full weight is given to the impulse (learning rate $\delta = 1$) that a clear case of reciprocity resembling TFT in a social dilemma context occurs. *Second*, the affective tie mechanism is backward-looking and automatic, whereas reciprocity models typically involve strategic forward-looking behavior and – in psychological game theory models – the incorporation of higher-order (un)kindness beliefs.[10] Nevertheless, through emotional mediation the mechanism makes long-term (un)kind behavior possible. *Third*, the GTV construct allows for a propagation of (positive or negative) caring based on interaction experiences with unrelated others. *Fourth*, because the tie mechanism does not necessarily require the tracking of recognizable others (which may be impossible or too demanding; see below Eq. (7)), it allows for "generalized reciprocity", that is, passing on prior behavior without any need of knowing from whom that behavior was received, which can be a cost-effective mechanism for increasing group cooperation in various contexts (Pfeiffer, Rutte, Killingback, Taborsky, & Bonhoeffer, 2005; Salazar, Shaw, Czekóová, Staněk, & Brázdil, 2022).[11]

5. Concluding discussion

This paper presents a novel uncertainty-based approach to why people – or organisms, more generally – automatically care for interaction counterparts: the informational affective tie mechanism (*i*ATM). It addresses all the issues raised with respect to the social ties model of Bault et al. (2017) in the Introduction. As a spin-off it also provides a dynamic endogenous and context-related representation of the well-known psychological construct of social value orientation (SVO).

[10] In these models reciprocity is either based on the expectation of sufficient reciprocity in return (think of Trivers' (1971) "reciprocal altruism"), on rational cooperation in finitely repeated games with incomplete information (Kreps, Milgrom, Roberts, & Wilson, 1982), or on the incorporation of (un)kindness beliefs into the utility function (Rabin, 1993; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006).

[11] As actions of different unrecognized interaction counterparts are likely to show greater variance, behavioral uncertainty increases (σ_ε^2 in the *i*ATM model; see Section 2, Module 1). Consequently, actions are less reliable and greater weight will be put on the prior estimate of counterpart's type (see Eqs. (2) – (4)).

Due to the ubiquity of uncertainty for organisms, the proposed *i*ATM model opens a perspective of multi-level bonding that bridges species and scientific disciplines.

Empirical support for the model comes, among others, from econometric evidence, the model's predictive power regarding various experimental game findings, and its relationship with a neural network model of maternal care (seen as primordial caregiving system), including oxytocin as (type-) signaling molecule.

Due to space constraints, a number of interesting additional implications of the mechanism cannot be discussed in full here (see the more extensive working paper, van Winden (2021)). To mention a few, consider first the possibility of tipping points through the build-up (or breakdown) of ties in affective networks facing a social dilemma. Think of a prisoner's dilemma, for instance. Notice that in that case cooperation (instead of defection) becomes the dominant choice once a sufficient positive weight is attached to the payoff of the interaction partner. Furthermore, note that the dyadic affective ties maintained by followers of a charismatic leader with that leader (emotional leadership) can help solve free-riding problems in large groups. It only requires that the leader is motivated (for intrinsic or extrinsic reasons) to call upon their tie-based intrinsic motivation to follow an exhortation by the leader to contribute (Loerakker & van Winden, 2017). Finally, notice that affective networks shed a new light on the definition of what an individual (as in methodological individualism) or an individual's identity actually is. 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 employed in economics. The *i*ATM model suggests a biopsychological extension of this standard approach by focusing on the role of affective ties in defining a person's "sense of self". Accordingly, an individual's identity is here proposed to comprise all agents – selves and others – one is intrinsically motivated to exert effort for (albeit to a different, attention and tie-value related extent). Consistent with this proposal is Hershfield's (2011) argument, regarding the effort of long-term planning, that what matters is an identity comprising both the current and the future self, where the degree of psychological connectedness between the two may vary (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 the previous Section). In fact, the *i*ATM model's dynamic view of identity provides a link with the biological concept of organismality (Queller & Strassmann, 2009; West & Kiers, 2009). The contextual nature of extended utility in the model, furthermore, seems consistent with the concept of "contextual organismality" proposed by Diaz-Muñoz et al. (2016).

Some issues left for future research are concluded with. First of all, further empirical investigation of the model is needed, in particular, regarding the modeling and specification of uncertainty. In this paper, the common procedure of using probability functions is followed, turning uncertainty into risk (see Section 2). This neglects the possibility of surprises in the form of unexpectedness, which can be important, for instance, if the environment profoundly changes. As noticed before, where relevant, the model can be extended to allow for strategic intertemporal decision-making (Bault et al., 2017; Loerakker et al., 2022). Also, the demarcation of contexts should be further investigated. In the literature on cooperation the pre-eminent context focused on is the prisoner's dilemma (or related games). Other contexts of interest, however, are 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 dominant-subordinate relationship dimension – get associated in memory (either separately or in some integrated form). Relatedly, further modeling and testing is required regarding 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. Another issue, finally, concerns the specification of the reference action in the stimulus (Eq. (1)). A natural candidate is the behavior expected from a counterpart that is neither friend nor foe (as assumed in Section 2). A more complicated model allowing for forward-looking behavior (see Bault et al., 2017) could incorporate behavior related to internalized norms.

References

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

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

Aimone, J., Ball, S., & King-Casas, B. (2015). The betrayal aversion elicitation task: an individual level betrayal aversion measure. *PLoS ONE (10)9*, e0137491, 1-12.

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

Ashar, Y.K., Andrews-Hanna, J.R., Dimidjian, S., & Wager, T.D. (2017). Empathic care and distress: predictive brain markers and dissociable brain systems. *Neuron, 94(6),* 1263-1273.

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

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

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

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

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

Bartels, D. M., & 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., & 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., & 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.

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

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.

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

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10,* 122-142.

Bethlehem, R. A. I., van Honk, J., Auyeung, B., & 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. & Feldman Barrett, L. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, *63*, 235–248.

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

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

Bosman, R., & van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal, 112,* 147-169.

Bosman, R., Hennig-Schmidt, H., & van Winden, F. (2017). Emotion at stake – The role of stake size and

emotions in a power-to-take game experiment in China with a comparison to Europe. *Games, 8,* 17.
Bosman, R., Sutter, M., & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology, 26,* 407-429.
Bowles, S., & Gintis, H. (2011). *A Cooperative Species.* Princeton: Princeton University Press.
Brandts, J., Riedl, A., & 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., & Van Volkenburgh, E. (2006). Plant neurobiology: an integrated view of plant signaling. *TRENDS in Plant Science*, *11(8)*, 413-419.
Brosnan, S. F., & de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*, 129–152.
Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, *6(147)*, 1-19.
Brucks, D., & von Bayern, A. M. P. (2020). Parrots voluntarily help each other to obtain food rewards. *Current Biology*, *30*, 1–6.
Burnham, T., McCabe, K., & Smith, V. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization, 43,* 57-73.
Camerer, C. F. (2003). Behavioral Game Theory. Princeton: Princeton University Press.
Chang, L.J., & Sanfey, A.G. (2009). Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining. *Frontiers in Behavioral Neuroscience*, *3,* 1-12.
Churchland, P. S., & Winkielman, P. (2012). Modulating social behavior with oxytocin: How does it work? What does it mean? *Hormones and Behavior, 61*, 392–399.
Cronin, K.A. (2012). Prosocial behaviour in animals: the Influence of of social relationships. *Animal Behaviour*, *84*, 1085-1093.
Daw, N. D. (2014). Advanced reinforcement learning. In P. W. Glimcher and E. Fehr (Eds.), *Neuroeconomics* (2nd ed.). Amsterdam: Elsevier.
Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, Supplement, 1218-1223.
Decety, J. & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience, 2*, 1–24.
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., & 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., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611-1618.
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., & Bronstein, J. L. (2016). Contextual organismality: Beyond pattern to process in the emergence of organisms. *Evolution*, *70(12)*, 2669–2677.
Douglas, A. J., & 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.
Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47(2),* 268-298.
Evers, E., de Vries, H., Spruijt, B. M., & 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., & 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., & Ridderinkhof, K. R. (2012). Neural correlates of dynamically evolving interpersonal ties predict prosocial behavior. *Frontiers in Neuroscience*, *6(28)*.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54(2),* 293-315.
Faraji, M., Preuschoff, K., & Gerstner, W. (2018). Balancing new against old information: the role of surprise in learning. *Neural Computation*, *30(1)*, 34-83.
Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137-140.
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.
Feldman, R., Monakhov, M., Maayan, P., & Ebstein, R.P. (2016). Oxytocin pathway genes: evolutionary ancient system impacting on human affiliation, sociality, and psychopathology. *Biological Psychiatry, 79(3),* 174-184.
Fellbaum, C. R., Gachomo, E. W., Beesetty, Y., Choudhari, S., Strahan, G. D., Pfeffer, P. E., Kiers, E. T., & Bücking, H. (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *PNAS*, *109(7)*, 2666-2671.
Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *PNAS*, *107(12)*, 5334–5338.
Fredrickson, B.L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19(3),* 313-332.
Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43(5),* 349–358.
Frijda, N. (2007). *The Laws of Emotion.* Mahwah, New Jersey: Erlbaum.
Friston, K. (2010) The free-energy principle: a unified brain theory? *Nature Reviews | Neuroscience 11,* 127-138.
Gabaix, X., & Laibson, D. (2017). Myopia and discounting. CEPR Discussion Paper 11914.
Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience and Biobehavioral Reviews*, *37*, 2608-2620.
Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics, 115*, 811–846.
Glimcher, P. W., & Fehr, E. (Eds.) (2014). *Neuroeconomics* (second edition). London: Academic Press.
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., & Pelphrey, K. A. (2013). Oxytocin enhances brain function in children with autism. *PNAS*, *110(52)*, 20953–20958.
Greiff, M., Ackermann, K. A., & 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., & 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.
Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3,* 367-388.
Guzmán, Y. F., Tronson, N. C., Jovasevic, V., Sato, K., Guedea, A. L., Mizukami, H., Nishimori, K., & Radulovic, J. (2013). Fear-enhancing effects of septal oxytocin receptors. *Nature Neuroscience*, *16*, 1185-1187.
Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & 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.
Ijzerman, H., & Denissen, J.J.A. (2019). Social value orientation and attachment: a replication and extension of Van Lange et al. (1997). *Royal Society Open Science, 6(4),* 181575.
Insel, T. R., & Young, L. J. (2001). The neurobiology of attachment. *Nature Reviews Neuroscience*, *2*, 129-136.
Johnson, N. D., & 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., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, *112(2)*, 375-404.
Kelly, A. M., & 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., & 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., & Bücking, H. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*, *333*(*6044*), 880-882.
Knight, F.H. (1921). Risk, Uncertainty and Profit. Boston: Houghton Mifflin.
Kramer, J., & Meunier J. (2016). Kin and multilevel selection in social evolution: a never-ending controversy? *F1000Research*, *5*, 1-13.
Kreps, D.M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory, 27(2),* 245-252.
Kummel, M., & Salant, S. W. (2006). The economics of mutualisms: optimal utilization of mycorrhizal mutualistic partners by plants. *Ecology*, *87(4)*, 892-902.
Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46(4)*, 352–367.
Liakoni, V., Modirshanechi, A., Gerstner, W., & Brea, J. (2021). Learning in volatile environments with the Bayes factor surprise. *Neural Computation*, *33(2),* 269-340.
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.
Loerakker, B., Bault, N., Hoyer, M., & van Winden, F. (2022). On the development of cooperative and antagonistic relationships in public good environments: a model-based experimental study. OSF Preprints. https://doi.org/10.31219/osf.io/wur7c.
Loerakker, B., & van Winden, F. (2017). Emotional leadership in an intergroup conflict game experiment. *Journal of Economic Psychology*, *63*, 143–167.
Loewenstein, G., O'Donoghue, T., & Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, *2*, 55-81.
Massen, J. J. M., Sterck, E. H. M., & de Vos, H. (2010). Close social associations in animals and humans: functions and mechanisms of friendship. *Behaviour*, *147*, 1379-1412.
Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.
Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio*, *3(2)*, 1-7.
Murphy, R.O., & Ackermann, K.A. (2014). Social value orientation: theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review, 18(1)*, 13–41.
Murphy, R.O., Ackermann, K.A., & Handgraaf, M.J.J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6(8),* 771–781.
Nelson, E. E., & Panksepp, J. (1998). Brain substrates of infant–mother attachment: contributions of opioids, oxytocin, and norepinephrine. *Neuroscience and Biobehavioral Reviews*, *22(3)*, 437–452.
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. (2020). *The Parental Brain.* Oxford: Oxford University Press.
Numan, M., & Young, L. J. (2016). Neural mechanisms of mother-infant bonding and pair bonding: similarities, differences, and broader implications. *Hormones and Behavior*, *77*, 98–112.
Olff, M., Frijling, J. L., Kubzansky, L. D., Bradley, B., Ellenbogen, M. A., Cardoso, C., Bartz, J. A., Yee, J. R., & 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.
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.
Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M., & Bonhoeffer, S. (2005). *Proceedings: Biological Sciences*, *272(1568)*, 1115-1120.
Pitcher, D., Japee, S., Rauth, L., & 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., & 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., & Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B*, *364*, 3143–3155.
Quintana, D. S., & Guastella, A. J. (2020). An allostatic theory of oxytocin. *Trends in Cognitive Sciences*, *24(7)*, 515-528.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83,* 1281–1302.
Salazar, M., Shaw, D.J., Czekóová, K., Staněk, R., & Brázdil, M. (2022). The role of generalised reciprocity and reciprocal tendencies in the emergence of cooperative group norms. *Journal of Economic Psychology*, *90,* 102520.
Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300,* 1755–1758.
Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.) (2001). *Appraisal Processes in Emotion.* Oxford: Oxford University Press.
Schino, G., & Aureli, F. (2009). Reciprocal altruism in primates: Partner choice, cognition, and emotions. *Advances in the Study of Behavior*, *39*, 45–69.
Schino, G., & Aureli, F. (2010). Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology, 19,* 130–135.
Schultz, W., Dayan, P., & Montague, R. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–99.
Seyfarth, R. M., & Cheney, D. L. (2012). The evolutionary origins of friendship. *Annual Review of Psychology*, *63*, 153–177.
Simms, E. L., & 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., & Oliveira, R. F. (2012). Arginine vasotocin regulation of interspecific cooperative behaviour in a cleaner fish. *PLoS ONE*, *7(7)*, e39583, 1-10.
Soutschek, A., Ruff, C. C., Strombach, T., Kalenscher, T., & Tobler, P. N. (2016). Brain stimulation reveals crucial role of overcoming self-centeredness in self-control. *Science Advances*, *2(10)*, e1600992.
Tennie, C., Jensen, K., & Call, J. (2016). The nature of prosociality in chimpanzees. *Nature Communications*, *7*, 13915, 1-8.
Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46(1)*, 35-57.
van Dijk, F., Sonnemans, J., & van Winden, F. (2002). Social ties in a public good experiment. *Journal of Public Economics*, *85*, 275–299.
van Dijk, F., & 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 longterm relationships. In R. Noë, J. A. R. A. M. van Hooff, & P. Hammerstein (Eds.), *Economics in Nature.* Cambridge: Cambridge University Press.
Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & 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., & Avilés, L. (2010). It takes grouping and cooperation to get sociality. *Journal of Theoretical Biology*, *264*, 1240–1253.
van Winden, F. (2021). The informational affective tie mechanism: on the role of uncertainty, context, and

attention in caring. Tinbergen Institute Discussion Paper TI 2021-012/I.
West, S. A., & Kiers, E. T. (2009). Evolution: what is an organism? *Current Biology*, *19(23)*, R1080-R1082.
West, S. A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. R. Soc. Lond.* B, *269*, 685-694.
Young, L. J,, & Wang, Z. (2004). The neurobiology of pair bonding. *Nature Neuroscience*, *7(10)*, 1048-54.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This is a thoroughly rewritten and abridged version of the Tinbergen Institute paper TI 2021-012/I, with the same title. A visiting professorship at the Department of Economics and Business of the University of Catania (September-November, 2021) and very helpful comments by two reviewers, as well as the editors, are gratefully acknowledged.

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:

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

for some δ_{ijt}, and some independent distributed noise term ξ_t, with variance σ^2_{ijt+1}. Multiplying both sides of (A1) by I'_{ijt} and taking expectations, gives: $E\left[\alpha'_{ijt+1} I'_{ijt}\right] = \delta_{ijt} E\left[I'^2_{ijt}\right]$; thus,

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

Next, taking the variance of both sides of (A1), gives: $\sigma^2_{ijt} = \delta^2_{ijt}\sigma^2_{I't} + \sigma^2_{ijt+1}$, with $\sigma^2_{I't} = \sigma^2_{ijt} + \sigma^2_{\varepsilon}$.

Using (A2), $\delta_{ijt}\sigma^2_{I't} = \sigma^2_{ijt}$, and, thus,

(A3) $\sigma^2_{ijt+1} = \sigma^2_{ijt} - \delta^2_{ijt}\sigma^2_{I't} = \sigma^2_{ijt} - \delta_{ijt}\sigma^2_{ijt} = \left(1-\delta_{ijt}\right)\sigma^2_{ijt}$.

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 $\delta_{ijt} = \sigma^2_{ijt} \,/\, \left(\sigma^2_{ijt} + \sigma^2_{\varepsilon}\right)$.