

Socializing Willpower: Resolve from the Outside In

By Stephen Setman and Daniel Kelly

Commentary on George Ainslie, “Willpower With and Without Effort” Forthcoming in *Behavioral and Brain Sciences*

Email: ssetman@purdue.edu and drkelly@purdue.edu

Website: <https://setman.carrd.co/> and <http://web.ics.purdue.edu/~drkelly/>

11-16-20

Abstract:

Ainslie’s account of willpower is conspicuously individualistic. Since other people, social influence, and culture appear only peripherally, it risks overlooking what may be resolve’s deeply social roots. We identify a general “outside-in” explanatory strategy suggested by a range of recent research into human cognitive evolution, and suggest how it might illuminate the origins and more social aspects of resolve.

Main Text:

Contemporary work on human evolution traces the ecological success of our species to sophisticated capacities for cooperation and culture (Boyd and Richerson 2005, Henrich 2015) and a suite of underlying psychological mechanisms selected to increase the efficacy with which humans interact with and learn from each other (Chudek and Henrich 2011, Sterelny 2012, Mathew and Perreault 2015, Laland 2017). This indicates that many distinctive forms of human cognition are fundamentally *social* (Kelly and Hoburg 2017).

Such capacities may be social even when they pretheoretically appear otherwise. For example, Mercier and Sperber (2011, 2017) argue against an orthodox view according to which reasoning initially evolved to enable individuals to deliberate and make better decisions, and only afterwards acquired social functions for e.g. facilitating interpersonal interactions and collective negotiations. Their “argumentative theory” offers an alternative picture, according to which reasoning is *primarily* social, and the mechanisms that underpin it initially evolved to perform the public functions of persuading others and assessing their attempts to persuade. Only once they emerged and established their social-oriented functions were those psychological mechanisms able to be turned back on the self, allowing reason to acquire the more private dimension that accompanied its new, interior, individual-oriented functions. Their theory also compels a shift from an individual to a socially-oriented evaluative perspective that reveals the mechanisms that guide reasoning as often performing their original public functions *well*, rather than merely being riddled with maladaptive biases causing them to do private reasoning poorly. Following Sterelny (2018), we will call this an instance of *outside-in explanation*.

Outside-in explanation is, if not widespread, becoming more common. Carruthers (2009, 2011) argues that our capacity to know the contents of our own minds is, perhaps counterintuitively, parasitic on mechanisms that originally evolved to know the minds of others. Introspection is a derivative, private usage of mindreading machinery that is primarily other-oriented, then reflexively turned back on oneself. McAdams and colleagues (McLean et al 2007, McAdams and McLean 2013, McAdams 2018) develop a similarly outside-in explanation of narrative identity. It sees an individual's creation and maintenance of their own narrative identity as a by-product of social mechanisms for mindreading, language, and storytelling. Others argue that features of human cognition like confirmation bias (Peters 2020) and over-imitation (Hoehl et al 2019) that look like design flaws from the perspective of individual rationality are actually adaptations selected to perform important social functions. Under headings like “relational” and “dialogical”, philosophers from a range of backgrounds continue exploring outside-in-type explanations of gender (Haslanger 2000, Witt 2011), personal identity (Lindeman 2014, Carr 2020), and the collaborative character of human agency (Doris 2015) and autonomy (Stoljar 2015).

Is resolve, as Ainslie depicts it, amenable to outside-in explanation? The prospects are promising (c.f. Shea et al 2014). Many of Ainslie's key ideas and theoretical resources—perspective taking, bargaining, prisoner's dilemmas, credibility, excuses—are most at home in discussions of sociality. Even mental time travel, foresight, and the general expansion of the time horizons to which human minds are sensitive has been linked to selection pressures generated by the demands of increasing social complexity (Donald 1991, 2006, Suddendorf and Corballis 2007).

Assuming this connection is not accidental, what might a more specific outside-in explanation of resolve look like? Here is one possible trajectory. The human species arrives and thrives on the strength of hypertrophied capacities for cooperation and culture, central to which are enhanced mechanisms for reputation tracking (Santos et al 2011), social prediction (Frith and Frith 2006), and normative forms of conformity and enforcement (Kelly and Setman 2020, Wu et al 2020). These initially outward-oriented social mechanisms—some of which may *themselves* have culturally evolved and been socially acquired (Heyes 2018)—are reoriented inward and turned back onto oneself, repurposed to help solve personal intertemporal dilemmas. When an individual learns the trick, they become capable of the self-directed and self-regulating (c.f. McGeer and Pettit 2002) form of prediction and enforcement Ainslie calls resolve.

Ainslie notes the affinity with oath taking (21), another manifestly social practice. Resolve and its (hypothesized) transition from outer to inner might be further illuminated by considering its similarities to *avowal* (c.f. Kelly forthcoming). This suggestion sees the dynamics of resolve as a mostly private, internalized mimicking of dynamics whose original form is found in public acts of endorsement. An individual can publicly affirm a norm they have adopted, or announce to others an intention to behave in a certain way, as a means to help themselves abide by it. With such acts of social signaling, they attempt (whether or not they explicitly understand what they are doing in these terms, (21)) to realign others' expectations of them and thereby summon a special kind of social

influence onto themselves. More specifically, they create new *reputational stakes* that are tied to the credibility and status they have among people whose good opinion they value or need.

Public avowal can thus reorganize incentives, much like resolve does on Ainslie's account. With avowal, however, an individual attempts to manage their own behavior by actually changing something external to their own mind, namely the kind of person *others* see them as, and the expectations others have of how they will act. In succeeding to change those expectations, they change their actual incentives, thereby exerting a socially-directed form of ecological control over their own behavior (Clark 2007, Holroyd and Kelly 2016). With resolve, a person changes not the external social world but their own internal perspective, broadening the incentives taken into consideration and bundling them in ways that change how they bear on an immediate choice via how that choice bears on their ability to continue *seeing themselves* as the kind of person they want to be. More of the process is internal, but the strategy and elements are suggestively similar.

Remaining steadfast and resolute, especially in the face of temptation, is often an intensely personal, internal struggle. Nevertheless, we think it plausible that the inner resources one draws on, both to make resolutions and to abide by them, have an external, social provenance.

References

- Carr, D.T. (2020). Personal identity is social identity. *Phenomenology and The Cognitive Sciences*, 1-11.
- Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *The Behavioral and Brain Sciences*, 32(2): 121-38; discussion 138-82.
- Carruthers, P. (2011). *The Opacity of the Mind: An Integrative Theory of Self-Knowledge*. New York: Oxford University Press.
- Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15, 218–226.
- Clark, A. (2007). Soft Selves and Ecological Control. In D. Spurrett, D. Ross, H. Kincaid & L. Stephens (Eds.) *Distributed Cognition and the Will*. Cambridge, MA: The MIT Press, 101–21.
- Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.
- Donald, M. (2007). The slow process: A hypothetical cognitive adaptation for distributed cognitive networks. *Journal of Physiology-Paris*, 101, 214-222.
- Doris, J. (2015) *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.
- Frith, C., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, 1079, 36-46.

- Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Noûs*, 34, 31-55.
- Henrich, J. (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.
- Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Cambridge, MA: Harvard University Press.
- Hoehl, S., Keupp, S., Schleihauf, H., McGuigan, N., Buttelmann, D., & Whiten, A. (2019). ‘Over-imitation’: A review and appraisal of a decade of research. *Developmental Review*, 51, 90-108.
- Holroyd, J. & Kelly, D. (2016). Implicit Bias, Character, and Control. In A. Masala & J. Webber (Eds.), *From Personality to Virtue: Essays in the Philosophy of Character* (pp. 106-133). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198746812.003.0006>
- Kelly, D. (forthcoming). Two Ways to Adopt a Norm: The (Moral?) Psychology of Avowal and Internalization. In M. Vargas & J. Doris (Eds.), *The Oxford Handbook of Moral Psychology*.
- Kelly, D., & Hoburg, P. (2017). A tale of two processes: On Joseph Henrich’s *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. *Philosophical Psychology*, 30(6), 832–848.
- Kelly, D. & Setman, S. (2020). The Psychology of Normative Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). URL = [<https://plato.stanford.edu/archives/fall2020/entries/psychology-normative-cognition/>](https://plato.stanford.edu/archives/fall2020/entries/psychology-normative-cognition/).
- Laland, K. (2017). *Darwin’s Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press.
- Lindemann, H. (2014). *Holding and Letting Go: The Social Practice of Personal Identities*. New York: Oxford University Press.
- Mathew, S., & Perreault, C. (2015). Behavioural variation in 172 small-scale societies indicates that social learning is the main mode of human adaptation. *Proceedings of the Royal Society B: Biological Sciences*, 282: 20150061.
- McAdams, D. (2018). “First we invented stories, then they changed us”: The Evolution of Narrative Identity. *Evolutionary Studies in Imaginative Culture*, 3(1): 1-18.
- McAdams, D., & McLean, K. (2013). Narrative Identity. *Current Directions in Psychological Science*, 22(3), 233–38. <https://doi.org/10.1177/0963721413475622>
- McGeer, V. & Pettit P. (2002). The Self-Regulating Mind. *Language & Communication*, 22, 281-299.

- McLean, K. C., Pasupathi, M., & Pals, J. (2007). Selves Creating Stories Creating Selves: A Process Model of Self-Development. *Personality and Social Psychology Review*, 11(3), 262–78.
<https://doi.org/10.1177/1088868307301034>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57 - 74.
- Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Peters, U. (2020). What Is the Function of Confirmation Bias? *Erkenntnis*, 1-26.
<https://doi.org/10.1007/s10670-020-00252-1>
- Richerson, P., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Santos, M.D., Rankin, D., & Wedekind, C. (2011). The evolution of punishment through reputation. *Proceedings of the Royal Society B: Biological Sciences*, 278, 371 - 377.
- Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18, 186 - 193.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: The MIT Press.
- Sterelny, K. (2018). “Why reason? Hugo Mercier's and Dan Sperber's *The Enigma of Reason: A New Theory of Human Understanding*,” *Mind & Language*, 33(5): 502-512.
<https://doi.org/10.1111/mila.12182>
- Stoljar, N. (2015). Feminist Perspectives on Autonomy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition). URL =
<https://plato.stanford.edu/archives/fall2015/entries/feminism-autonomy/>.
- Suddendorf, T., & Corballis, M. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *The Behavioral and brain sciences*, 30 3, 299-313; discussion 313-51 .
- Witt, C. (2011). *The Metaphysics of Gender*. New York: Oxford University Press.
- Wu, J., Balliet, D., Peperkoorn, L.S., Romano, A., & Lange, P.A. (2019). Cooperation in Groups of Different Sizes: The Effects of Punishment and Reputation-Based Partner Choice. *Frontiers in Psychology*, 10, 2956.